High hopes for low-growing corn plants p. 364

Menopause in chimpanzees pp. 368 & 416

Wavelike exciton flow p. 438

Science

$15
27 OCTOBER 2023
SPECIAL ISSUE
science.org
AAAS

THE HYPOTHALAMUS

Coordinating basic survival functions p. 386

REVIEWS

The structural and functional complexity of the integrative hypothalamus p. 388

Neurocircuits for motivation p. 394

Hypothalamic control of innate social behaviors p. 399

Sleep and the hypothalamus p. 405

RELATED ITEM

SCIENCE REVIEW BY J. C. BRÜNING AND H. FENSELAU 10.1126/SCIENCE.ABL7398

SMALL AND MIGHTY

THE HYPOTHALAMUS

By **Mattia Maroso** *and* **Peter Stern**

If you pause for a second and think about the activities that occupy most of your day, presumably sleeping, eating, and engaging in social interactions are among the first that come to your mind. Perhaps surprisingly, a small area buried deep inside the brain, called the hypothalamus, is responsible for coordinating neuronal signals related to these activities. By controlling the homeostasis of the neuroendocrine, limbic, and autonomic nervous systems, the hypothalamus is a key brain region for many physiological and pathological processes. Despite its small size, the hypothalamus has a complex cellular organization and circuitry that determine its structural and functional organization. It is composed of 11 nuclei grouped by their location and has vast, mostly bidirectional connections with many neuronal and endocrine systems.

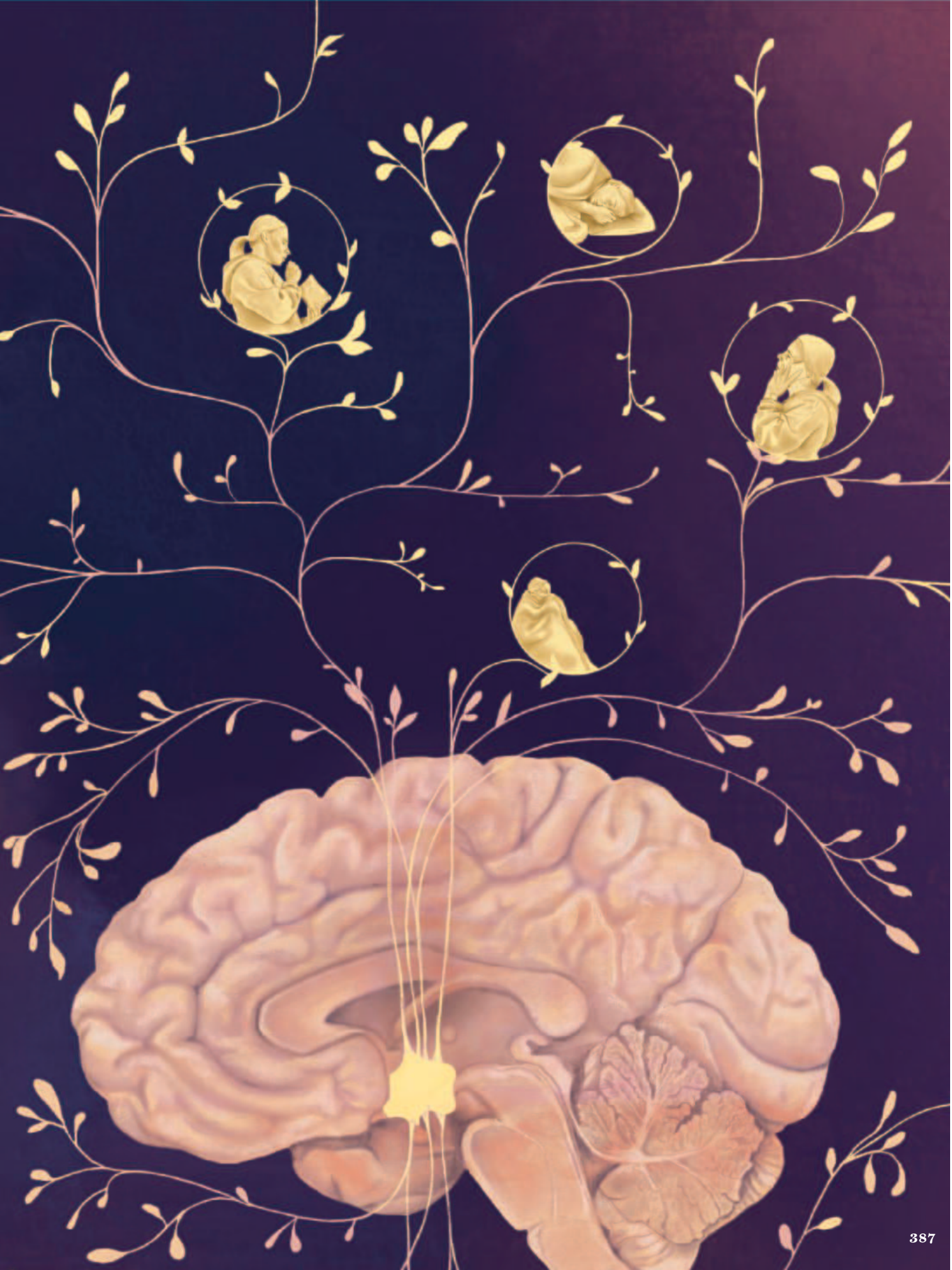
In this special issue, we highlight the main role that the hypothalamus plays in integrating the many body functions required for the maintenance of homeostasis. If that wasn't enough, it also modulates social interactions, including affective, sexual, and aggressive behaviors. Yet another still not completely understood function of the hypothalamus is the way it controls sleep. Finally, to avoid missing the forest for the trees, we have to take a broader view of the neural circuits that mediate motivated behaviors and their interactions with the dopaminergic system to reinforce ongoing or planned actions to fulfill motivational demands.

The hypothalamus may be small, but it has a colossal impact on a wide range of behaviors. Once again, the power of David over Goliath resides in organization and action, not size.

10.1126/science.adl3437

The image depicts the main functions regulated by the hypothalamus. Temperature, sleep, eating, and social interactions are all controlled by this small, but mighty, area.

ILLUSTRATION: GRACE RUSSELL



REVIEW

The structural and functional complexity of the integrative hypothalamus

Harmony Fong[1,2,3]†, Jing Zheng[1,2,3]†, Deborah Kurrasch[1,2,3]*

The hypothalamus ("hypo" meaning below, and "thalamus" meaning bed) consists of regulatory circuits that support basic life functions that ensure survival. Sitting at the interface between peripheral, environmental, and neural inputs, the hypothalamus integrates these sensory inputs to influence a range of physiologies and behaviors. Unlike the neocortex, in which a stereotyped cytoarchitecture mediates complex functions across a comparatively small number of neuronal fates, the hypothalamus comprises upwards of thousands of distinct cell types that form redundant yet functionally discrete circuits. With single-cell RNA sequencing studies revealing further cellular heterogeneity and modern photonic tools enabling high-resolution dissection of complex circuitry, a new era of hypothalamic mapping has begun. Here, we provide a general overview of mammalian hypothalamic organization, development, and connectivity to help welcome newcomers into this exciting field.

The hypothalamus makes up just 2% of brain volume but is essential for fundamental life processes and survival. The term hypothalamus was first introduced by Wilhelm His in 1893 and was further annotated by Santiago Ramón y Cajal in the 1890s, Albrecht von Haller in the 1900s, and Ernst and Berta Scharrer in the 1920s. For centuries, the hypothalamus has fascinated scientists for its nodal position and dense nuclear clustering of neurons, and this allure only increased as early lesioning studies revealed the outsized role of the hypothalamus in a range of physiologies and behaviors. Today, the hypothalamus is known to receive, process, and integrate sensory inputs to drive bidirectional communication with a range of behavioral, autonomic, and endocrine pathways. The development and interaction of the circuits by which the hypothalamus maintains homeostatic set points, overcomes stressors, and drives behaviors are continually being defined.

The intricate organization of the hypothalamus

The hypothalamus resides in the ventral-most region of the forebrain and drives diverse processes through a complex cytoarchitecture localized around the third ventricle (*1*, *2*) (Fig. 1A). Along the rostral-caudal axis, the hypothalamus can be divided into the preoptic, anterior, tuberal, and mammillary regions (Fig. 1, B to E), each of which is further partitioned into lateral, medial, and periventricular zones (*3*). These hypothalamic regions are populated by a patchwork of neurons that form discrete three-dimensional (3D) clusters, or nuclei, which contrasts with the relatively flat laminar structure of the cerebral cortex. Most rostrally, the preoptic area (POA) includes the median (MnPO), medial (MPO), ventrolateral (VLPO), anteroventral periventricular (AVPV), and periventricular (PVpo) preoptic nuclei. The anterior region is composed of the anterior hypothalamus (AH) and supraoptic (SON), suprachiasmatic (SCN), and paraventricular (PVN) nuclei. The median eminence (ME) and pituitary stalk lie ventral to the tuberal hypothalamus, which contains the loosely defined lateral hypothalamic area (LH) as well as the dorsomedial (DMH), ventromedial (VMH), arcuate (Arc), and tuberal (TuN) hypothalamic nuclei. Finally, the premammillary (PM), supramammillary (SuM), tuberomammillary (TM), and posterior hypothalamic (PH) nuclei make up the mammillary hypothalamus.

The mapping of functional roles and accompanying pathways onto individual nuclei has revealed extensive overlap and high coordination (*1*, *2*) (Fig. 1F). For example, various aspects of feeding, energy balance, and thermoregulation are strongly regulated by neurons residing in the Arc, VMH, DMH, LH, and TuN within the tuberal hypothalamus, but they also loop in circuits within the MnPO, VLPO, PVN, AH, and PM. Similarly, corticotropin-releasing hormone (CRH) neurosecretory cells in the PVN are central to stress responses and integrate inputs from the MPO, DMH, LH, and PH (*4*). Further, sleep-wake is orchestrated by the SCN—the principal circadian pacemaker—in conjunction with the VLPO, MnPO, PVN, DMH, VMH, LH, TM, and SuM (*5*). Although the interconnectedness of spatially segregated nuclei is a core feature of hypothalamic function, more recently this concept has evolved to appreciate the concerted contributions of extrahypothalamic circuits as well. For example, the hypothalamus facilitates circuitry for reward, salivation, smell, and hunger alongside the physical act of foraging for food, which provides a framework for studying complex behaviors mediated by the hypothalamus, such as feeding.

Some hypothalamic nuclei are known to be sex dimorphic. The MPO is one of the best-known sex-biased brain regions, with male sexual behaviors in rats influenced by a greater vol[illegible] and neuronal number compared with tha[illegible] females (*6*). Notably, this anatomical size difference is also observed in the third interstitial nucleus of the anterior hypothalamus (INAH-3), the presumptive homologous cell group in humans (*7*). In mice, distinct subpopulations of tyrosine hydroxylase– and progesterone receptor–expressing neurons in the AVPV and the VMH, respectively, are sexually dimorphic and mediate a range of sexual, parental, and aggressive behaviors (*8*, *9*), whereas female-biased differences in pro-opiomelanocortin (POMC) neuronal numbers and activity in the Arc have been reported to influence energy balance (*10*). High-resolution mapping of hypothalamic circuitry mediating sex-specific behaviors is an emerging area of research (*11*).

Extraordinary diversity of hypothalamic cell types

To better understand the cellular basis of hypothalamic functions, attention has largely focused on the resident neurons that are staggering in diversity. Elegant biochemical studies in the 1950s identified various neuropeptidergic neurons across the hypothalamus (*12*), with the true extent of hypothalamic neuron heterogeneity now being realized with the emergence of single-cell omics, particularly single-cell RNA sequencing (scRNA-seq). Although hypothalamic neurons historically have been classified on the basis of neuropeptide phenotype or fast neurotransmitter identity, scRNA-seq–defined transcriptional signatures highlight the complexity of hypothalamic neuronal populations. For example, some neuropeptide transcripts are preferentially expressed in glutamatergic neurons (e.g., *Oxt*, *Grp*, and *Ucn3*), whereas others are found in γ-aminobutyric acid–expressing (GABAergic) neurons (e.g., *Npy* and *Agrp*) or even both (e.g., *Sst* and *Crh*) (*13*). Concomitantly, the broad categorization of excitatory glutamatergic or inhibitory GABAergic populations can be further stratified into subpopulations that express a combination of neuropeptides and transcription factors (*14*–*16*), such as *Hcrt* and *Pnoc* distinguishing clusters of *Lhx9*+ putative glutamatergic neurons within the LH (*15*). Seventeen hypothalamic scRNA-seq datasets have been integrated into a consolidated and comprehensive mouse hypothalamic cell atlas, serving as a valuable resource for studying hypothalamic function with unprecedented cell subtype–specific precision (*17*).

The challenge now is to translate this transcriptional profiling into higher-resolution insights of the cell type-function relationships. For instance, a subset of *Gal*+ and *Avpr1a*+ neurons in the POA are associated with enriched cFos expression after pup exposure, which suggests a role in parenting behavior (*18*); subtypes of orexigenic agouti-related peptide (AgRP) neurons in the Arc exhibit differential transcriptional responses to fasting, potentially representing

[1]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. [2]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada. [3]Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada.
*Corresponding author. Email: kurrasch@ucalgary.ca
†These authors contributed equally to this work.

discrete, energy state–responsive populations (*19*); and molecularly distinct clusters of neurons in the mammillary bodies are not only spatially segregated but also project to different regions of the anterior thalamus as perhaps parts of parallel memory subcircuits (*20*). To build on these largely correlative links and establish function, fine dissection of the component parts of diverse hypothalamic circuits will require combining newly identified subtype-specific markers with targeted genetic tools. For example, chemogenetic activation of LH *Sst*+ neurons increases locomotor activity (*21*), whereas activating VMH *Esr1*+ neurons elicits both locomotion and heat production (*22*), which highlights that circuits mediating movement behaviors are dependent on context. In addition, chemogenetic and optogenetic manipulations of broad VMH *Esr1*+ neurons reveal this population to be necessary and sufficient for female aggression (*23*), whereas stereotaxic injection of small interfering RNAs (siRNAs) against *Rprm*, which is expressed in a subset of VMH *Esr1*+ neurons, increases body temperature in female mice, consistent with a sex-specific thermoregulatory role for this neuronal subpopulation (*22*). These studies underscore the arduous task of fine mapping the hypothalamic cell types and circuits that control complex behaviors.

Despite a neuron-centric view of hypothalamic cell type diversity, the hypothalamus also contains heterogeneous glial populations, defined originally in classical morphological studies and more recently by multi-omics approaches. These glial cells, including astrocytes, oligodendrocyte precursor cells, oligodendrocytes, microglia, ependymal cells, and tanycytes, are not simply structural support but are active participants in modulating hypothalamic circuits (*24*). Astrocytes in the SCN, for instance, show rhythmic expression of clock genes and influence circadian locomotor behavior (*25*), and microglial activity in the developing POA is required for mating behaviors (*26*). Glial contributions to the hypothalamic control of energy balance are also extensive, ranging from leptin response modulation by astrocytes and oligodendrocyte precursor cells; to melanin-concentrating hormone (MCH)–responsive cilia beating by ependymal cells, which may in turn facilitate volume transmission of MCH through the cerebrospinal fluid; to nutrient sensing, metabolite transport, and neurogenesis by multifunctional tanycytes (*24*, *27*). Altogether, mounting evidence indicates that different classes of glia play essential roles across neuroendocrine networks, with further studies needed.

Developmental programs of the hypothalamus

Given the diversity of cellular phenotypes found in the hypothalamus, an understanding of when and how these cell identities are acquired has lagged. According to the columnar model of early hypothalamic organization, and with the exception of the telencephalon-derived POA, the hypothalamus begins to develop from the ventral diencephalon at approximately embryonic day (E) 8.5 in mice or gestational week (GW) 5 in humans, with both extrinsic and intrinsic factors directing regional patterning (*13*). Sonic hedgehog (Shh) drives the expression of the classic hypothalamic marker *Nkx2.1* and is required to induce a hypothalamic fate in diencephalic progenitors, whereas Bmp and Wnt signaling play important roles in establishing the dorsal-ventral and rostral-caudal axes, respectively (*28*). After early patterning, the developing hypothalamus is populated by neurons, then glia, following the typical pattern for other neural tube–derived brain regions: Radial glia divide at the ventricular zone, giving rise to fate-committed neuronal and glial precursors that undergo radial migration into the parenchyma (*2*). In the developing mouse and human hypothalamus, hypothalamic radial glia lining the third ventricle can divide asymmetrically to generate mantle radial glial cells, which serve as intermediate progenitor cells (IPCs) located in the mantle zone (*29*) (Fig. 2A). Notably, live-cell imaging of mantle radial glia reveals that in addition to undergoing asymmetric neurogenic divisions, they can also produce cells expressing the progenitor marker SOX2 (*29*). Some of these daughter cells have basal processes and likely function to maintain the mantle radial glia pool, whereas others that lack processes are speculated to increase the diversity of hypothalamic progenitors and/or to contribute to gliogenesis (*29*).

Early hypothalamic precursors are heterogeneous (*30*), supportive of a cascade diversification model for hypothalamic neurogenesis. Unlike

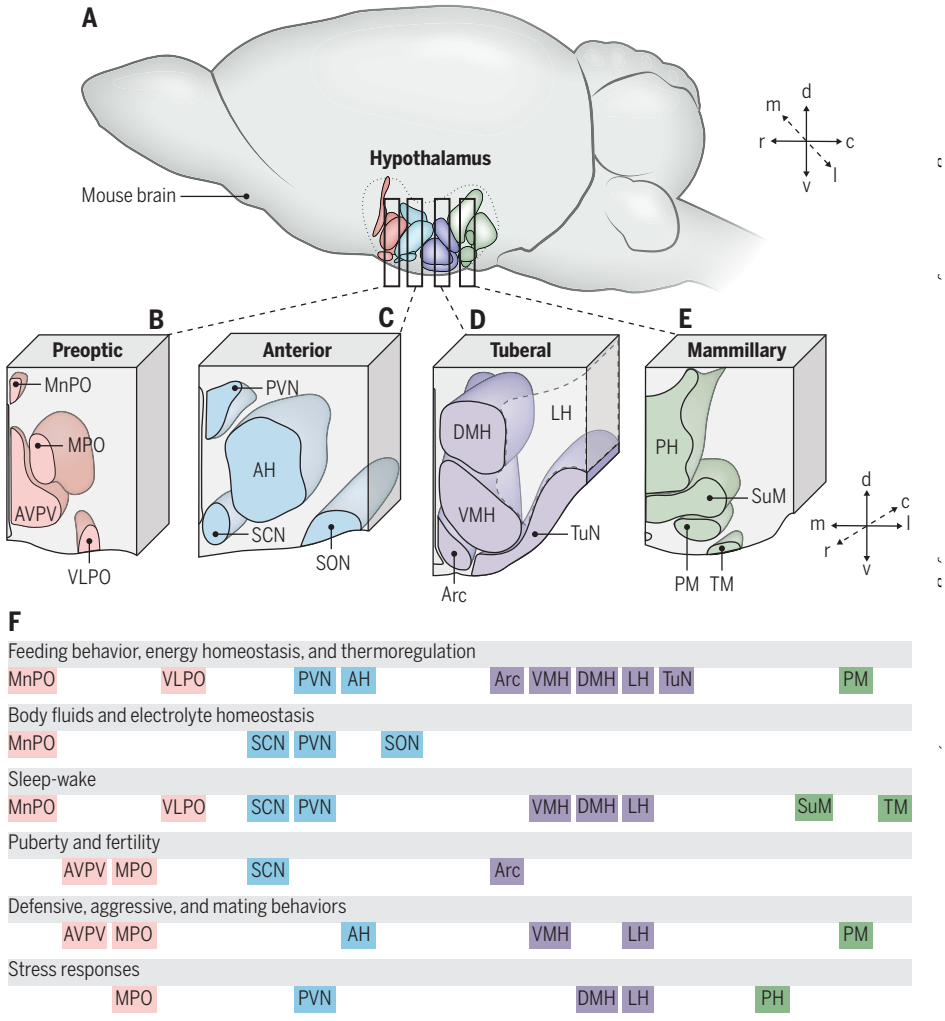


Fig. 1. A schematic representation of hypothalamic subdivisions in the mouse brain, highlighting individual nuclei and associated physiologies. (**A**) The hypothalamus resides at the base of the brain, with its nuclei forming a complex 3D patchwork around the third ventricle across four rostral-caudal regions. (**B** to **E**) Representative coronal sections of the preoptic (B), anterior (C), tuberal (D), and mammillary (E) regions demonstrate the diversity and intercalation of hypothalamic nuclei. (**F**) The hypothalamic control of diverse physiologies and behaviors critical to survival involves coordinated activity across multiple nuclei.

CREDIT: N. BURGESS/*SCIENCE*, BASED ON HARMONY FONG, JING ZHENG, AND DEBORAH KURRASCH

the fate-predetermined and stochastic models of the cortex and the retina, respectively, the cascade diversification model posits that the myriad neuronal subtypes required for hypothalamic function are generated in a stepwise amplification of radial glial cells, to intermediate progenitors, to neurons (*31*). In brief, hypothalamic radial glia give rise to *Ascl1*+ IPCs, competent to generate both glutamatergic and GABAergic neurons, and *Neurog2*+ IPCs, which primarily produce glutamatergic neurons (*31*) (Fig. 2A). The postmitotic neuronal progeny of these complementary pools of intermediate progenitors ultimately populate diverse nuclei. In their nascent state, the coexpression of multiple peptides in a given neuronal subtype, along with pseudotime analyses, suggest that several neuronal phenotypes can emerge from a common, initially ambiguous transcriptional state (*31*), which points to hypothalamic complexity from its origin. Notably, in the Arc, fate-mapping experiments reveal that both somatostatin (SST) and AgRP neurons are derived from SST-expressing nascent neurons (*31*), and distinct subpopulations of neurons defined by combinatorial *Tbx3*, *Otp*, and *Dlx1/2/5/6* expression are generated from a common multipotent progenitor domain in a mosaic and synchronous fashion (*32*), providing evidence for this fate diversification.

As in other areas of the brain, neuronal and glial cells in the hypothalamus emerge during specific temporal windows (Fig. 2B). Hypothalamic neurogenesis occurs at approximately E10.5 to E16.5 in mice (*33*) and GW8 to GW10 in humans (*34*), with neuronal precursors specified by the combinatorial expression of transcription factors and their corresponding regulons (*15*, *16*, *31*). In mice, neuronal fate commitment is marked by the *Dlx1/2/5/6* and *Sox11/12* regulons (*16*), the former of which also characterizes neurogenic progenitors in the developing human hypothalamus (*34*). Further region- and nucleus-specific identity is conferred by additional gene regulatory networks, such as *Lhx5* for the mammillary hypothalamus and *Foxa1* for the SuM (*15*), and neuronal generation proceeds in an outside-in, rostral-caudal pattern (*32*, *34*). Comparatively less is known about gliogenesis, but the process initiates upon the decline of neurogenesis (*16*, *34*) under the control of the *Hes5*, *Sox9*, and *Nfia* regulons (*16*). Tanycytes are among the first-born glial cells, with most generated between E13 and E15 (*31*), followed by the onset of oligodendrocyte lineage cells (E13.5 or GW10), astrocytes (E15.5 or GW15) (*34*, *35*), and finally ependymal cells (E17.5 or GW18) (*16*, *34*) (Fig. 2B). Notably, oligodendrocyte development occurs relatively early in humans compared with mice, with in silico cell-cell communication prediction suggesting a potential human-specific role for early oligodendrocyte lineage cells in facilitating neuronal survival and maturation through Wnt and integrin signaling (*34*). The role of glia in neuronal maturation and circuitry establishment continues to emerge as the genetic programs governing hypothalamic development are revealed.

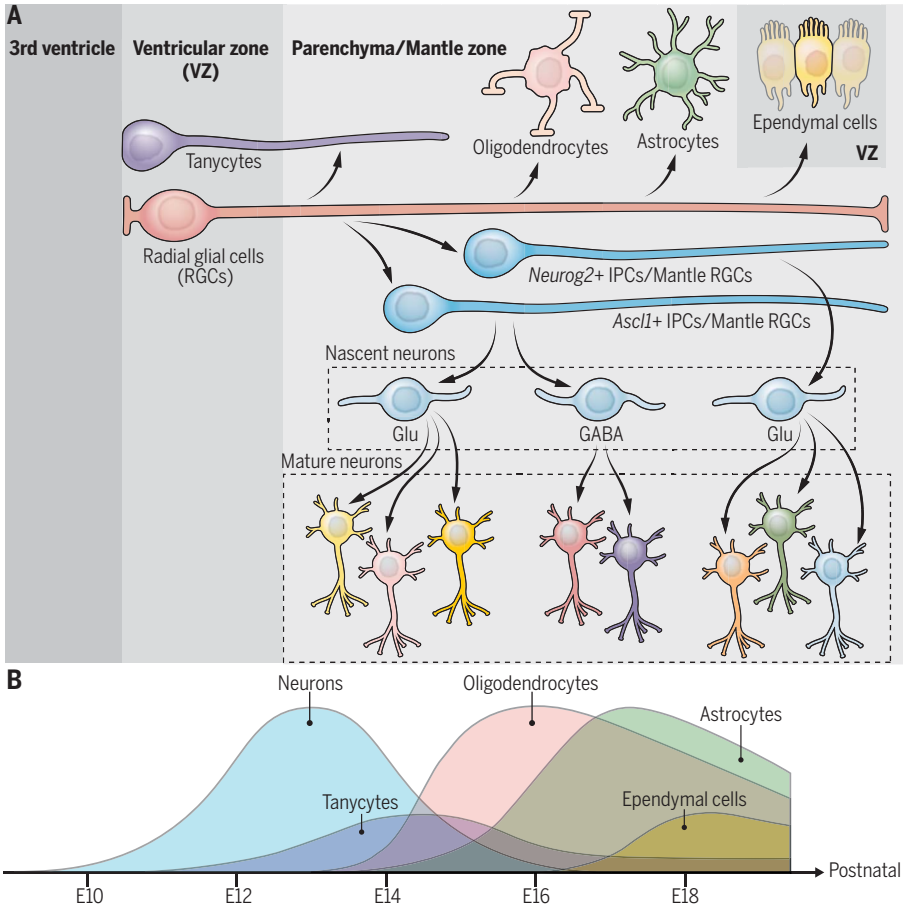


Fig. 2. Developmental programs of the mouse hypothalamus. (**A**) Radial glial cells (RGCs) divide at the ventricular zone (VZ) and give rise to both glia (tanycytes, oligodendrocytes, and ventricular zone–residing ependymal cells) and neurons. Hypothalamic neurogenesis is proposed to occur through a cascade diversification model in which ventricle-adjacent radial glia generate *Neurog2*+ and *Ascl1*+ intermediate progenitors located in the mantle zone. These intermediate progenitors then generate nascent glutamatergic or GABAergic neurons that ultimately differentiate into the range of neuronal subtypes found in the mature hypothalamus. (**B**) Schematic representation of the timing of neurogenesis and gliogenesis in the developing hypothalamus. Glu, glutamatergic; GABA, GABAergic.

Environmental influences on hypothalamic development

Curiously, the developing hypothalamus is influenced by the external environment, changing programs in response to the cues perceived to define the environment in which it will live (Fig. 3A). This concept of fetal programming or developmental origins of heath and disease (DOHaD) is not new, but it does seem especially pronounced in the murine fetal hypothalamus, perhaps reflecting the eventual central role of the hypothalamus as a sensing brain region responsible for adjusting a range of homeostatic and behavioral systems to match the current environment.

Accumulating evidence illustrates the permanent effects of environmental inputs during development on proper hypothalamic function later in life. In both human and animal models, a maternal high-fat diet can induce metabolic disorders in the offspring by altering Arc neuropeptides, signaling, neuroinflammation, and/or epigenetics (*36*). In mice, maternal cold stress causes hypothalamic fetal microglia to sex-specifically disrupt oxytocin neuronal numbers (*37*), and maternal fever (through exogenous immune activation) accelerates the onset of puberty with accompanying changes in hypothalamic *Kiss1*, *Tac2*, and *Kiss1r* gene expression (*38*). Furthermore, anthropological chemicals, especially endocrine-disrupting chemicals that mimic hormones, such as the xenoestrogen bisphenol A (BPA), can perturb developmental hypothalamic programs. Maternal exposure to environmentally relevant levels of BPA can cause precocious neurogenesis in the SCN of embryonic mice, leading to a concomitant disruption of circadian rhythms (*39*) as well as changes in the development of

CREDIT: N. RIRKESS/SCIENCE; BASED ON HARMONY FONG, JING ZHENG, AND DEBORAH KURRASCH

kisspeptin and gonadotropin-releasing hormone (GnRH) neurons, resulting in fertility issues in females (*40*). Beyond bisphenols, many studies have shown that exposure to air pollution particle matter, diesel exhaust, polychlorinated biphenyls, and nicotine can interfere with hypothalamic neuro- and gliogenesis (*41*). The true impact of environmental inputs on human hypothalamic development is less defined, with large cohort studies needed to further delineate their influence.

Hypothalamic connectivity that drives physiology and behaviors

Given that hypothalamic development is a protracted series of steps whereby a neuron born at a specific time migrates to the correct location and forms appropriate synaptic connections, it is perhaps not surprising that complex developmental programs progress into intricate circuit formation. As a neural integration center, the hypothalamus connects to the endocrine system via the pituitary and the circumventricular organs (CVOs) as well as to other neural centers through intra- and extrahypothalamic circuits that control behaviors (Fig. 3). The sophistication of these connections is central to the myriad physiologies and behaviors that the hypothalamus controls.

Neuroendocrine connectivity

The hypothalamus connects to the anterior and posterior lobes of the pituitary via the infundibulum (also known as the pituitary stalk). Briefly, magnocellular neurons that originate in the PVN and SON send axons that travel through the infundibulum and terminate in the posterior pituitary, where they secrete arginine-vasopressin and oxytocin to be released into peripheral circulation. The connection to the anterior pituitary is less direct, with neurons in the POA and other regions secreting hypothalamic-releasing hormones (e.g., GnRH and CRH) into the ME, where the hypophyseal-portal vasculature transports them to the anterior pituitary to stimulate the synthesis and secretion of pituitary hormones (e.g., luteinizing hormone, follicle-stimulating hormone, and adrenocorticotropic hormone) that then act on peripheral endocrine organs (*42*). Notably, the release of these neuropeptides or hormones is either contributed by multiple nuclei (i.e., arginine-vasopressin and oxytocin by the PVN and SON) or is the result of interactions among multiple nuclear clusters (e.g., the release of follicle-stimulating hormone by the interactions among AVPV, POA, and Arc), which emphasizes the extensive cooperation between hypothalamic nuclei (Fig. 1F).

The neuroendocrine hypothalamus also relies on highly permeable brain regions that reside outside the blood-brain barrier to sense peripheral cues, referred to as the CVOs: the organum vasculosus of the lamina terminalis (OVLT), the subfornical organ (SFO), and the ME (Fig. 3A). The OVLT resides in the anterior wall of the third ventricle and detects circulating chemokines and cytokines to initiate body temperature changes as well as monitor solute levels to drive regulation of water balance and thirst (*43*). The SFO is positioned in the dorsal surface of the third ventricle and is the primary sensor of circulating angiotensin II to communicate with brain centers that regulate blood pressure, sympathetic outflow, and neuroendocrine factors (*43*). The ventral-residing ME's primary function is to provide a portal for the secretion of hypothalamic-releasing hormones into the anterior pituitary, but its capillary bed is also in direct contact with hypothalamic ventricular cells and Arc neurons, which suggests a role in sensing peripheral signals (*43*). Recently, a SCN-OVLT vascular portal was identified that retrogradely transports arginine-vasopressin from the hypothalamus to the OVLT for secretion into broader circulation (*44*), thereby creating a second portal system outside the hypophyseal vasculature, which leads to intriguing questions as to whether hypothalamic neurons can secrete neuropeptides for peripheral circulation independent of the pituitary.

Hypothalamic neural connectivity

Individual hypothalamic nuclei are highly interconnected to facilitate the coordinated regulation

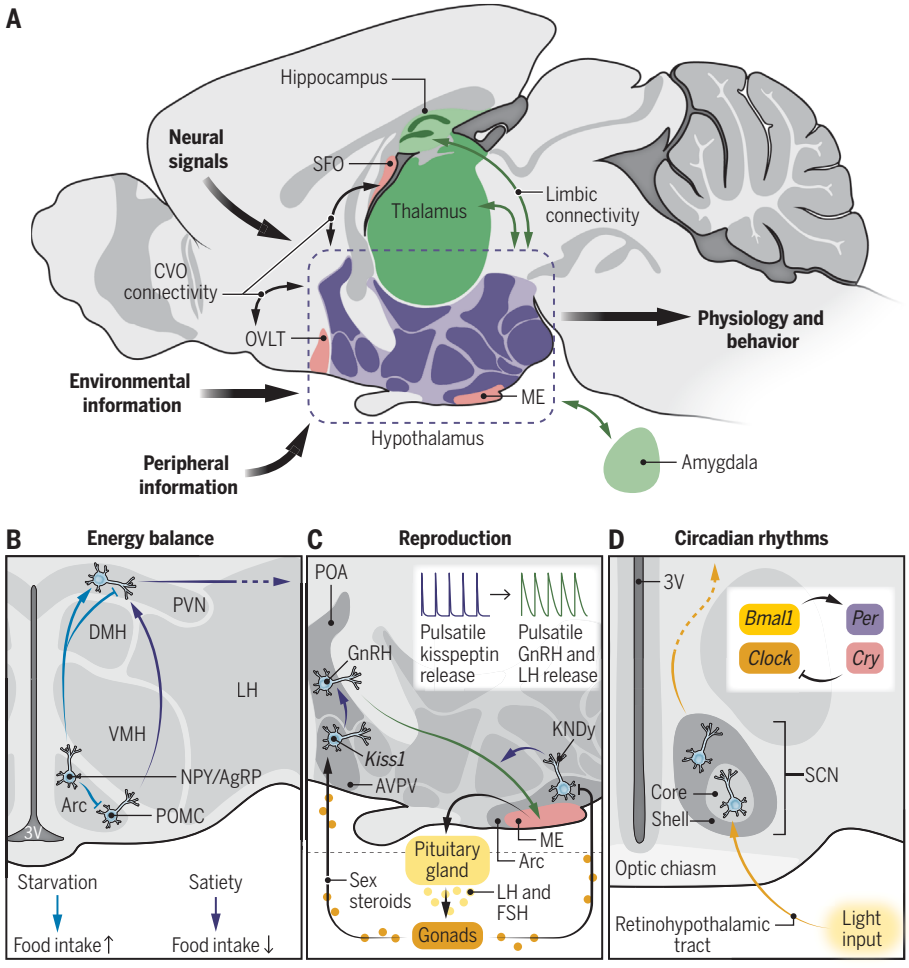


Fig. 3. The hypothalamus as an integration center driving key physiologies. (**A**) The hypothalamus receives and integrates environmental information, peripheral information, and neural signals to regulate physiologies and behaviors via connections to the endocrine system and intra- and extrahypothalamic neural centers. (**B**) After food intake, POMC neurons in the Arc that project to PVN neurons increase satiety, thus decreasing food intake (purple). During fasting, NPY/AgRP-expressing neurons in the Arc that project to the PVN increase food intake while also decreasing satiety by inhibiting (i) POMC neurons directly and (ii) the anorexigenic POMC inputs on PVN neurons (blue). (**C**) *Kiss1* neurons in the AVPV together with KNDy neurons in the Arc act on GnRH neurons in the POA to control the pulsatile release of GnRH, luteinizing hormone (LH), and follicle-stimulating hormone (FSH) to regulate female fertility. (**D**) Responding to light input from the retinal-hypothalamic tract, the SCN maintains circadian rhythms through the autoregulatory transcriptional-translational feedback loops of clock genes, such as *Bmal1*, *Clock*, *Period1,2,3*, and *Cry1,2*. 3V, third ventricle.

CREDIT: N. BURGESS/*SCIENCE*, BASED ON HARMONY FONG, JING ZHENG, AND DEBORAH KURRASCH

of energy balance, reproduction, sleep-wake, stress responses, thermoregulation, osmoregulation, and blood pressure. Moreover, the hypothalamus's central positioning within the brain facilitates broader neural connectiveness, such as through the dorsal longitudinal fasciculus to the brainstem, across the medial forebrain bundle to the neocortex, via the stria terminalis to the amygdala, and within the mammillothalamic tract to the thalamus, among others. Each of these pathways contributes to a range of neural processes, making the hypothalamus an important contributor to a multitude of behaviors (*45*). Herein, we provide a high-level overview of the hypothalamus's role in the limbic system plus three well-established circuits to illustrate the interconnectiveness of hypothalamic nuclei: energy balance, reproduction, and circadian rhythms.

Alongside the amygdala, thalamus, and hippocampus, the hypothalamus is a key component of the limbic system. The limbic circuitry serves as a crucial interface between cognitive functions, emotions, and adaptational responses and allows for the formation of memories, learning from experiences, and behavioral adaptations in response to the environment. The major hypothalamic-limbic connections include the hippocampal-centered circuits that connect via the fornix to regulate the anterograde episodic memory and the amygdala-centered circuits interconnected with the stria terminalis and thalamus that are considered fundamental for the processing of mood, affect, and emotions (*42*, *45*) (Fig. 3A).

Centrally, the hypothalamus relies on a complex interplay of hypothalamic circuitry to regulate energy balance. Simplistically, the drive to forage is mediated by neurons coexpressing orexigenic neuropeptide Y (NPY) and AgRP (NPY/AgRP) in the Arc that activate melanocortin hormone– and orexin-expressing neurons in the LH that increase food intake, whereas satiety is primarily controlled by POMC neurons in the Arc that project to PVN neurons to release α-melanocyte–stimulating hormone and decrease food intake. Second-order neurons in the PVN, VMH, DMH, and LH also send additional intra- and extrahypothalamic projections that help regulate feeding (*46*). Notably, when energy levels are low, NPY/AgRP-expressing neurons can directly inhibit POMC neurons to effectively block satiation (*47*) (Fig. 3B).

Alongside energy balance, the hypothalamus plays a fundamental role in regulating reproduction. Prenatally, GnRH neurons migrate from the olfactory placode to the POA, where they remain quiescent until acted upon at puberty by pulsatile kisspeptin release from Arc neurons that coexpress kisspeptin, neurokinin B, and dynorphin (KNDy) and drives synchronized GnRH secretion (*48*) (Fig. 3C). KNDy neurons together with the kisspeptin (*Kiss1*) neurons in AVPV control the feedback loop that regulates this pulsatile gonadotropin release and induces sex steroid synthesis and secretion from the gonads as well as gametogenesis (*49*). KNDy neurons are involved in the negative feedback of sex steroids, whereas *Kiss1* neurons in the female AVPV are involved in the positive feedback of hormones leading to the preovulatory luteinizing hormone surges (Fig. 3C). Given that reproduction is an energy-demanding event, a tight relationship between metabolic status and reproduction exists, with melanocortin neurons and kisspeptin neurons in the Arc influencing each other (*50*), although the full mechanisms are still emerging.

The hypothalamus is the central circadian pacemaker in mammals, with the SCN receiving light input from the retinohypothalamic tract to entrain the environmental light-dark cycle (Fig. 3D). Mechanistically, neurons of the SCN generate autonomous cellular circadian oscillations through the autoregulatory transcriptional-translational feedback loops of clock genes, such as *Bmal1*, *Clock*, *Period1,2,3*, and *Cry1,2*, which then regulate large gene networks to control synchrony of cellular activities (*51*). These SCN neurons regulate circadian rhythms by interacting with other wakefulness- and sleep-promoting circuits in the POA, DMH, PVN, and LH, releasing transforming growth factor–α (TGF-α) and prokineticin 2 and sensing changes in body temperature (*5*), for example. The LH is especially crucial in controlling sleep-wake cycles, with orexin neurons promoting wakefulness and maintained arousal during the day (*5*).

Emerging circuits of the hypothalamus

Advancements in cellular techniques, such as cell type–specific viral transfections, antero- and retrograde tracing, and chemogenetic and optogenetic manipulations, facilitate a higher-resolution probing of hypothalamic connectivity, leading to the identification of emerging circuits. Recently, the PVN and the LH were revealed to help regulate pain, with parvocellular oxytocin neurons in the PVN simultaneously projecting to magnocellular oxytocin neurons in the SON and deep-layer neurons in the spinal cord to suppress nociception and promote analgesia (*52*). Moreover, activation of PVN neurons that project to the neocortex reduces pain sensitization and anxiety-like behaviors (*53*), whereas LH orexin neurons that project to the dorsal horn of the spinal cord mediate antinociceptive effects (*54*). These findings implicate the hypothalamus as an underappreciated region involved in the regulation of autonomic and motivational responses to pain.

Moreover, the hypothalamus is emerging as a neural integration center for gut-derived hormones and microbiota metabolites. Serotonin, cholecystokinin, and peptide YY hormones secreted by the gut can inhibit the activity of hunger-promoting hypothalamic AgRP neurons to induce satiety (*55*), whereas secretin released from colon endocrine cells indirectly activates neurons in the nucleus of the solitary tract that project to the PVN to modulate sodium homeostasis (*56*). In addition, muropeptides derived from gut bacteria can be transferred to the hypothalamus, where they act on GABAergic neurons to regulate food intake and body temperature (*57*). Less understood is the influence of the maternal microbiome on developmental programs of the hypothalamus, although maternal treatment with the so-called good bacterium, *Lactobacillus*, corrects a gut microbial imbalance from a maternal high-fat diet, restoring oxytocin levels in the developing PVN and reversing social deficits in offspring (*58*). Mapping the hypothalamic circuitry responsive to the gut microenvironment will require a multidisciplinary approach that combines microbiome profiling and multi-omics integration with rodent gnotobiotic and behavioral studies.

The hypothalamus as a debated adult neurogenic center

A less-recognized output of the hypothalamus is its influence on adult neurogenesis in the two key neurogenic centers: the neocortex and the hippocampus (Fig. 4A). In response to metabolic cues, POMC neurons in the Arc send projections to the ventricular-subventricular zone of the neocortex and regulate neural stem cell proliferation that drives adult neurogenesis of deep granule neurons (*59*). Similarly, optogenetic stimulation of hypothalamic SuM neurons robustly promotes adult neuronal birth in the hippocampal dentate gyrus, leading to increased production of adult-born neurons and improved memory performance (*60*). These studies raise the compelling possibility that the hypothalamus serves as a conduit, translating sensory inputs into outcomes in adult neurogenic niches.

At the same time, an ongoing debate continues as to whether the hypothalamus itself is a bona fide adult neurogenic center. The ventricular zone of the adult hypothalamus has neurogenic potential, with low levels of adult-born hypothalamic neurons found in the POA and the tuberal hypothalamus. Tanycytes, specialized glia that reside in this niche, have a radial glia–like morphology and transcriptomic signature (*61*), which suggests a shared role in cell genesis (Fig. 4B). Short-term exposure to a high-fat diet induces hypothalamic ventricular cells to proliferate and give rise to neurons (*62*), with genetic fate mapping linking adult neurogenesis to tanycytes in the ME (*63*). Evidence is not conclusive as to whether these newly born neurons are functionally integrated into hypothalamic circuits, although these newly born tanycyte-derived neurons display some electrophysiological properties of functional neurons (*64*), and transplanted immature hypothalamic neurons reconstitute leptin responsiveness in obese *db/db* mice (*65*), demonstrating the feasibility of adult integration. Given the low numbers of hypothalamic adult-born neurons relative

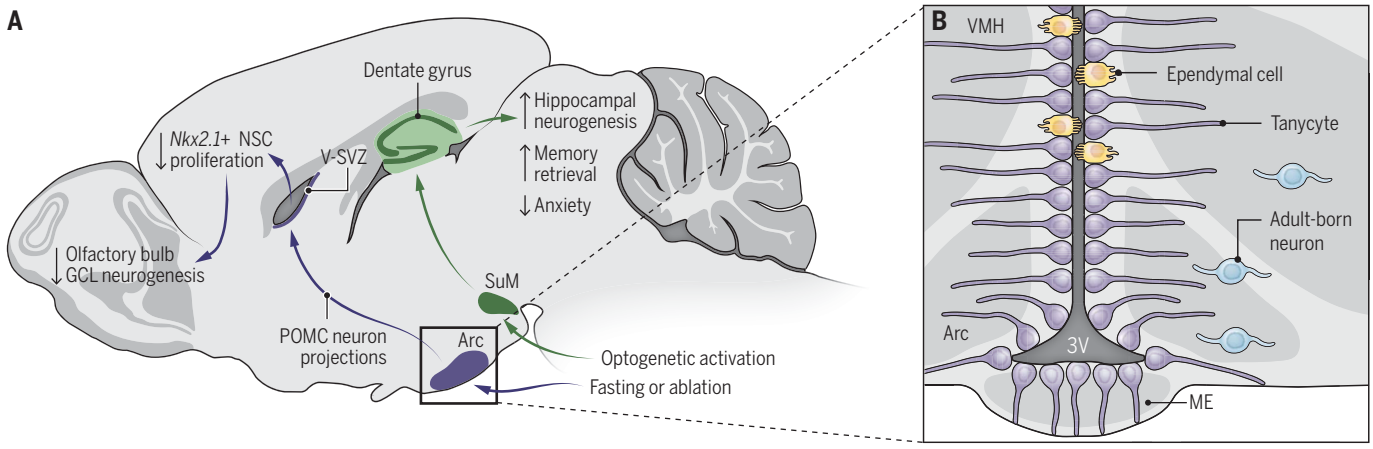


Fig. 4. Adult neurogenesis mediated by the hypothalamus. (**A**) In response to environmental cues, projections from POMC Arc neurons to the ventricular-subventricular zone (V-SVZ) regulate neural stem cell (NSC) proliferation, which then modulates adult neurogenesis in the deep granule cell layer (GCL) of the olfactory bulb. Optogenetic activation of SuM neurons increases adult neurogenesis in the hippocampal dentate gyrus, which improves memory performance after putative circuit integration. (**B**) Radial glia–like tanycytes are thought to contribute to adult neurogenesis in the Arc and VMH, among other hypothalamic regions.

to those produced by the neocortical and hippocampal niches (*62*), questions continue regarding the functional relevance of these adult hypothalamic neurons.

Future directions

The hypothalamus is an ancient brain region that predates the evolution of the vertebrate brain, with hypothalamic neurosecretory cells found in insects and annelids (*66*). Comparative studies show species-specific hypothalamic programs that are either gained or lost during evolution, which raises compelling questions about the developmental and functional relationships between the hypothalamus and specialized areas of the mammalian brain, especially the neocortex. For example, the hypothalamus is more complex in cytoarchitecture and cell diversity compared with the neocortex, where the columnar units are relatively consistent across the cortical plate despite mediating an extensive range of cognitive functions. This observation perhaps reflects an evolutionary simplification of the potentially surplus networks used by the hypothalamus or, alternatively, reveals the necessary redundancy in circuitry needed to ensure survival. Relatedly, the hypothalamus and neocortex use different strategies to yield neuronal diversification, which may further reflect an adoption of simplified programs in the neocortex. Indeed, some hypothalamic cell types common across vertebrates, such as ciliated ependymal serotonergic cells, are absent in the mammalian hypothalamus (*67*), but related homologs are found in the lateral ventricles, which raises questions as to whether some primitive hypothalamic functions have been assumed by other brain regions. Molecular and genetic tools, such as spatial transcriptomics and cell type–specific Cre lines, can provide insights into the relationship between hypothalamic cell types and the functional roles of related cells in other brain regions, especially the neocortex.

Historically, hypothalamic neurons have received the lion's share of research attention, but today, neuron-to-glia, glia-to-neuron, and glia-to-glia communication is more appreciated. Glial cells in the ventricular zone not only give rise to all cell types in the hypothalamus but also sense peripheral cues via the CVOs and transfer these signals to nearby neurons. Moreover, parenchymal glia, consisting of microglia, astrocytes, and cells of oligodendrocyte lineages, interact with neurons to control synaptic transmission, among other roles. Chronic high-fat diet in mice, for instance, can cause astrogliosis and microglia activation, leading to neuroinflammatory effects on nearby hypothalamic neurons. Although most hypothalamic axons are not myelinated, many input projections are, which makes defects in myelination potentially relevant to hypothalamic diseases, which remain poorly characterized. As the mapping of hypothalamic connectivity continues, the influence of neural glia on these circuits must also be examined.

Modern molecular tools have generated a wealth of data that enable the cataloging of hypothalamic cell types, which now must be interrogated for their roles in development and function. To do so, specialized Cre lines that target increasingly discrete subpopulations are needed. For instance, *Sst*+ neurons in the LH were recently stratified into four molecularly distinct populations (*21*), potentially each requiring their own Cre line to enable functional characterization. Furthermore, Cre drivers can be used in combination with chemogenetic and optogenetic lines to activate or inhibit neuronal populations, such as with diptheria toxin to ablate specific cell types, stereotaxic injections of siRNAs and/or short hairpin RNAs (shRNAs) to knock down target genes, or floxed reporters to study lineages. Furthermore, dead CRISPR-Cas9 enzymes (dCas9s) are increasingly used to activate or inhibit targeted transcription factors to study the combinatorial code driving developmental programs (*68*). Finally, the field must continue to pursue human model systems, including hypothalamic primary and organoid cultures enriched with various combinations of hypothalamic nuclei, to confirm developmental and network findings (where feasible) observed in rodents. In conclusion, deciphering the fundamental mechanisms of hypothalamic development and circuitry establishment is critical to our understanding of physiological homeostasis and behaviors and should facilitate the discovery of the underlying etiology of many disease states.

REFERENCES AND NOTES

1. C. B. Saper, B. B. Lowell, *Curr. Biol.* **24**, R1111–R1116 (2014).
2. Y. Xie, R. I. Dorsky, *Development* **144**, 1588–1599 (2017).
3. E. A. Markakis, *Front. Neuroendocrinol.* **23**, 257–291 (2002).
4. J. P. Herman *et al.*, *Compr. Physiol.* **6**, 603–621 (2016).
5. T. E. Scammell, E. Arrigoni, J. O. Lipton, *Neuron* **93**, 747–765 (2017).
6. R. A. Gorski, R. E. Harlan, C. D. Jacobson, J. E. Shryne, A. M. Southam, *J. Comp. Neurol.* **193**, 529–539 (1980).
7. L. S. Allen, M. Hines, J. E. Shryne, R. A. Gorski, *J. Neurosci.* **9**, 497–506 (1989).
8. N. Scott, M. Prigge, O. Yizhar, T. Kimchi, *Nature* **525**, 519–522 (2015).
9. C. F. Yang *et al.*, *Cell* **153**, 896–909 (2013).
10. C. Wang *et al.*, *Nat. Commun.* **9**, 1544 (2018).
11. D. W. Bayless *et al.*, *Cell* **186**, 3862–3881.e28 (2023).
12. R. Guillemin, *Science* **202**, 390–402 (1978).
13. M. Benevento, T. Hökfelt, T. Harkany, *Nat. Rev. Neurosci.* **23**, 611–627 (2022).
14. R. Chen, X. Wu, L. Jiang, Y. Zhang, *Cell Rep.* **18**, 3227–3241 (2017).
15. D. W. Kim *et al.*, *Nat. Commun.* **11**, 4360 (2020).
16. R. A. Romanov *et al.*, *Nature* **582**, 246–252 (2020).
17. L. Steuernagel *et al.*, *Nat. Metab.* **4**, 1402–1419 (2022).
18. J. R. Moffitt *et al.*, *Science* **362**, eaau5324 (2018).
19. J. N. Campbell *et al.*, *Nat. Neurosci.* **20**, 484–496 (2017).
20. L. E. Mickelsen *et al.*, *eLife* **9**, e58901 (2020).
21. L. E. Mickelsen *et al.*, *Nat. Neurosci.* **22**, 642–656 (2019).
22. J. E. van Veen *et al.*, *Nat. Metab.* **2**, 351–363 (2020).
23. K. Hashikawa *et al.*, *Nat. Neurosci.* **20**, 1580–1590 (2017).
24. J. Clasadonte, V. Prevot, *Nat. Rev. Endocrinol.* **14**, 25–44 (2018).
25. O. Barca-Mayo *et al.*, *Nat. Commun.* **8**, 14336 (2017).

CREDIT: N. BURGESS/*SCIENCE*, BASED ON HARMONY FONG, JING ZHENG, AND DEBORAH KURRASCH

26. K. M. Lenz, B. M. Nugent, R. Haliyur, M. M. McCarthy, *J. Neurosci.* **33**, 2761–2772 (2013).
27. S. Nampoothiri, R. Nogueiras, M. Schwaninger, V. Prevot, *Nat. Metab.* **4**, 813–825 (2022).
28. S. Burbridge, I. Stewart, M. Placzek, *Compr. Physiol.* **6**, 623–643 (2016).
29. X. Zhou *et al*., *Nat. Commun.* **11**, 4063 (2020).
30. D. W. Kim *et al*., *Cell Rep.* **38**, 110251 (2022).
31. Y. H. Zhang *et al*., *Cell Stem Cell* **28**, 1483–1499.e8 (2021).
32. T. Ma, S. Z. H. Wong, B. Lee, G. L. Ming, H. Song, *Neuron* **109**, 1150–1167.e6 (2021).
33. J. Altman, S. A. Bayer, *Adv. Anat. Embryol. Cell Biol.* **100**, 1–178 (1986).
34. X. Zhou *et al*., *Cell Stem Cell* **29**, 328–343.e5 (2022).
35. C. M. Marsters *et al*., *Neural Dev.* **11**, 20 (2016).
36. B. Harmancıoğlu, S. Kabaran, *Front. Genet.* **14**, 1158089 (2023).
37. J. M. Rosin, S. Sinha, J. Biernaskie, D. M. Kurrasch, *Dev. Cell* **56**, 1326–1345.e6 (2021).
38. X. Zhao, M. Erickson, R. Mohammed, A. C. Kentner, *Dev. Psychobiol.* **64**, e22278 (2022).
39. D. Nesan, K. M. Feighan, M. C. Antle, D. M. Kurrasch, *Sci. Adv.* **7**, eabd1159 (2021).
40. C. Pivonello *et al*., *Reprod. Biol. Endocrinol.* **18**, 22 (2020).
41. L. Koshko, S. Scofield, G. Mor, M. Sadagurski, *Front. Endocrinol.* **13**, 938094 (2022).
42. H. L. Müller *et al*., *Nat. Rev. Dis. Primers* **8**, 24 (2022).
43. G. I. Uwaifo, *The Human Hypothalamus: Anatomy, Dysfunction and Disease Management* (Springer, 2020).
44. R. Silver, Y. Yao, R. K. Roy, J. E. Stern, *J. Neuroendocrinol.* **35**, e13245 (2023).
45. P. Y. Risold, R. H. Thompson, L. W. Swanson, *Brain Res. Rev.* **24**, 197–254 (1997).
46. S. G. Bouret, in *Appetite and Food Intake: Central Control*, R. Harris, Ed. (CRC Press, ed. 2, 2017), pp. 135–154.
47. K. Timper, J. C. Brüning, *Dis. Model. Mech.* **10**, 679–689 (2017).
48. Y. Uenoyama, N. Inoue, S. Nakamura, H. Tsukamura, *Front. Endocrinol.* **10**, 312 (2019).
49. V. M. Navarro, M. Tena-Sempere, *Nat. Rev. Endocrinol.* **8**, 40–53 (2011).
50. V. M. Navarro, *Nat. Rev. Endocrinol.* **16**, 407–420 (2020).
51. J. S. Takahashi, *Nat. Rev. Genet.* **18**, 164–179 (2017).
52. M. Eliava *et al*., *Neuron* **89**, 1291–1304 (2016).
53. X. H. Li *et al*., *Cell Rep.* **36**, 109411 (2021).
54. Y. Jeong, J. E. Holden, *Neuroscience* **159**, 1414–1421 (2009).
55. L. R. Beutler *et al*., *Neuron* **96**, 461–475.e5 (2017).
56. Y. Liu *et al*., *Sci. Adv.* **9**, eadd5330 (2023).
57. I. Gabanyi *et al*., *Science* **376**, eabj3986 (2022).
58. S. A. Buffington *et al*., *Cell* **165**, 1762–1775 (2016).
59. A. Paul, Z. Chaker, F. Doetsch, *Science* **356**, 1383–1386 (2017).
60. Y. D. Li *et al*., *Nat. Neurosci.* **25**, 630–645 (2022).
61. T. Goodman, M. K. Hajihosseini, *Front. Neurosci.* **9**, 387 (2015).
62. S. Yoo, S. Blackshaw, *Prog. Neurobiol.* **170**, 53–66 (2018).
63. D. A. Lee *et al*., *Nat. Neurosci.* **15**, 700–702 (2012).
64. S. Yoo *et al*., *Sci. Adv.* **7**, eabg3777 (2021).
65. A. Czupryn *et al*., *Science* **334**, 1133–1137 (2011).
66. L. A. Lemaire, C. Cao, P. H. Yoon, J. Long, M. Levine, *Sci. Adv.* **7**, eabf7452 (2021).
67. T. K. Sato *et al*., *Neuron* **43**, 527–537 (2004).
68. C. T. Breunig *et al*., *PLOS ONE* **13**, e0196015 (2018).

ACKNOWLEDGMENTS

We thank Kurrasch laboratory members for stimulating discussions and V. Navarro for helpful insights. We apologize to those whose studies were not cited owing to space limitations. **Funding:** Research into hypothalamic development and function is funded by the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council of Canada (NSERC) (to D.K.), a CIHR postdoctoral fellowship to J.Z. (MFE-183073), and a NSERC graduate student fellowship to H.F. (PGSD-569191-2022). **Author contributions:** H.F. and J.Z. contributed equally and have the right to list their name first on their CVs. **Competing interests:** D.K. is the cofounder of Path Therapeutics, which is focused on the discovery of antiseizure medications and is not in conflict with the topics discussed herein. The other authors declare no competing interests.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 1 August 2023; accepted 28 September 2023
10.1126/science.adh8488

REVIEW

Neurocircuits for motivation

Garret D. Stuber[1,2,3]*

The nervous system coordinates various motivated behaviors such as feeding, drinking, and escape to promote survival and evolutionary fitness. Although the precise behavioral repertoires required for distinct motivated behaviors are diverse, common features such as approach or avoidance suggest that common brain substrates are required for a wide range of motivated behaviors. In this Review, I describe a framework by which neural circuits specified for some innate drives regulate the activity of ventral tegmental area (VTA) dopamine neurons to reinforce ongoing or planned actions to fulfill motivational demands. This framework may explain why signaling from VTA dopamine neurons is ubiquitously involved in many types of diverse volitional motivated actions, as well as how sensory and interoceptive cues can initiate specific goal-directed actions.

The scientific study of motivation and its essential biological substrates boasts a long and rich history rooted in both philosophy and science (*1*). Work carried out in the 20th century such as Pavlov's discovery of conditioning (*2*) and Skinner's research on reinforcement (*3*) laid the behavioral foundation for the study of motivation. These and other concepts led to the identification of specific "neural centers," including parts of the hypothalamus, that could produce or suppress specific motivated actions (*4*, *5*). Although we now appreciate that animal behavior encompasses a diverse spectrum of physiological and emotional components that are both innate and learned, we can simplify the motivating factors that guide our actions into two opposing primordial goals: efficiently obtaining rewards and avoiding threats or their proxies. In this Review, I operationally define motivation as measurable approach, avoidance, or action in response to specific sensory stimuli to fulfill a biological need. In a broad sense, biological needs encompass natural behaviors such as feeding, drinking, sleep, reproduction, safety, and territorial defense. Effort, on the other hand, requires some measure of the amount of work, force, or other physical exertion that an organism is willing to perform to receive or avoid some exogenous stimulus. Motivation does not necessarily require effort but is often highly correlated with it, with notable exceptions such as impulsive or habitual behaviors. However, a mismatch between motivation and effort expended to obtain or avoid something generally indicates a disrupted behavioral process found in several conditions, ranging from addiction to depression (*6*).

To start dissecting the biological basis of motivation, we need to consider how an organism's nervous system interacts with its environment to catalyze essential biological processes such as learning and reinforcement, emotion generation, physiological homeostasis, and ultimately goal-directed motor output. This task may seem daunting, given the complexity of any single process and the intricacy of animal behavior. However, research in rodents over largely the past decade has identified brain circuits and cell types that participate in innate or instinctive behaviors such as feeding, drinking, and avoidance of pain. This emerging understanding reveals a cohesive neurocircuit framework in which primary sensory and interoceptive stimuli are directly and indirectly conveyed to subcortical systems to direct adaptive goal-directed motivated behavioral output.

In this neurocircuit framework (Fig. 1), salient features of sensory information are first detected by genetically defined sensory neurons that innervate the periphery. These systems convey information about the animal's environment to evolutionary ancient brain regions such as the brainstem and hypothalamus as well as to primary and secondary sensory cortices. Lower-dimensional features of this sensory information are then routed to cortico-limbic regions where learning causal associations between environmental features and internal state occurs. These learning- and prediction-associated regions include the amygdala, hippocampus, and frontal cortices. Networks within these structures further process and aggregate sensory and learned information to guide subsequent adaptation and decision-making on the basis of current needs, environment, and experience. Glutamatergic output from this learning and prediction network targets subcortical structures, such as the hypothalamus, extended amygdala, and ventral striatum. They are essential for instinctive behavioral programs related to feeding, drinking, aggression, reproduction, and escape. Neurons within these subcortical subregions are genetically and functionally diverse, but many key molecular markers reliably mark discrete cell populations and circuits. Moreover, these circuits are directly and indirectly regulated by biological signals, including circulating hormones released from the periphery. Hormonal signaling within molecularly defined circuits further shapes t activity dynamics and molecular states to a incoming signals from the learning and sensory layers to fully engage instinct layer networks that are particularly critical for neural systems for social motivation, such as mating (*7*–*9*), aggression (*10*, *11*), and parenting behaviors (*12*–*14*). Neuronal output from the ventral striatum, extended amygdala, and hypothalamus primarily targets midbrain and hindbrain neurons, including those in the ventral tegmental area (VTA), where information associated with motivating stimuli converges to regulate the release of dopamine in the forebrain.

VTA circuits and cell types for motivated behavior

At the heart of motivational circuits are the VTA dopamine neurons, which send dense projections to the nucleus accumbens (NAc) of the ventral striatum to produce motivational gain control and feedback to drive the formation of reproducible actions (Fig. 2). Dopamine signaling drives reinforcement (*15*, *16*), produces incentive salience (*17*), encodes learning- related prediction errors signals (*18*, *19*), and conveys causal associations (*20*). Individual VTA dopamine neurons can also encode distinct aspects of sensory, motor, and cognitive variables related to motivation (*21*) as well as gastrointestinal signals related to feeding and drinking (*22*). Optogenetic stimulation of VTA dopamine neurons produces a generic motivation signal that can reinforce even arbitrary actions (*16*, *23*–*25*). Because of their relatively low number in the brain, and stereotypical response patterns, VTA dopamine neurons likely receive reduced forms of information arriving from multiple competing sources. Dopamine neurons require synaptic inputs from both local and distal cells to regulate their activity. This creates a temporally dynamic system to reinforce actions associated with a particular motivated behavior state in downstream basal ganglia circuits (*25*). Besides ventral striatal projecting dopamine neurons, dopamine neurons that target the prefrontal cortex, amygdala, and other structures likely support distinct functions (*26*, *27*).

VTA γ-aminobutyric acid (GABA)–releasing neurons are molecularly distinct (*28*) and are the principal regulator of VTA dopamine neurons to inhibit their activity, reduce reward consumption (*29*), and produce generalized aversion (*30*). Furthermore, VTA GABAergic neurons provide a subtractive signal to VTA dopamine neurons that inhibits them when reward is expected, causally contributing to prediction error calculations (*31*). Most primary afferents to the VTA target both GABAergic and dopaminergic neurons (*32*, *33*), where monosynaptic inputs to these cells can encode aspects of prediction error signals, but do not recapitulate full prediction-error signals observed

[1]Center for the Neurobiology of Addiction, Pain, and Emotion, University of Washington, Seattle, WA 98195, USA. [2]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA 98195, USA. [3]Department of Pharmacology, University of Washington, Seattle, WA 98195, USA.
*Corresponding author. Email: gstuber@uw.edu

in VTA dopaminergic neurons (*34*). Excitatory synaptic inputs from the lateral habenula (LHb) predominantly serve as a source of excitatory drive to the VTA GABAergic neurons, promoting aversion or avoidance (*24*, *35*). LHb neurons that project to VTA function to convey aversive stimuli information, which can immediately lead to reductions in dopaminergic activity and cause animals to disengage from ongoing behaviors or promote escape behavior. Reductions in dopamine neuron activity frequently correlate with aversion or prediction errors; however, there is increased dopamine release in the medial shell of the NAc and the medial prefrontal cortex after aversive stimuli presentation, further highlighting the diversity of dopamine neurons in the VTA (*26*, *27*). LHb neurons themselves receive direct glutamatergic input from the lateral hypothalamic area (LHA) (*36*–*38*), ventral pallidum (*39*), extended amygdala (*40*), and anterior hypothalamic regions such as the lateral preoptic area (*41*).

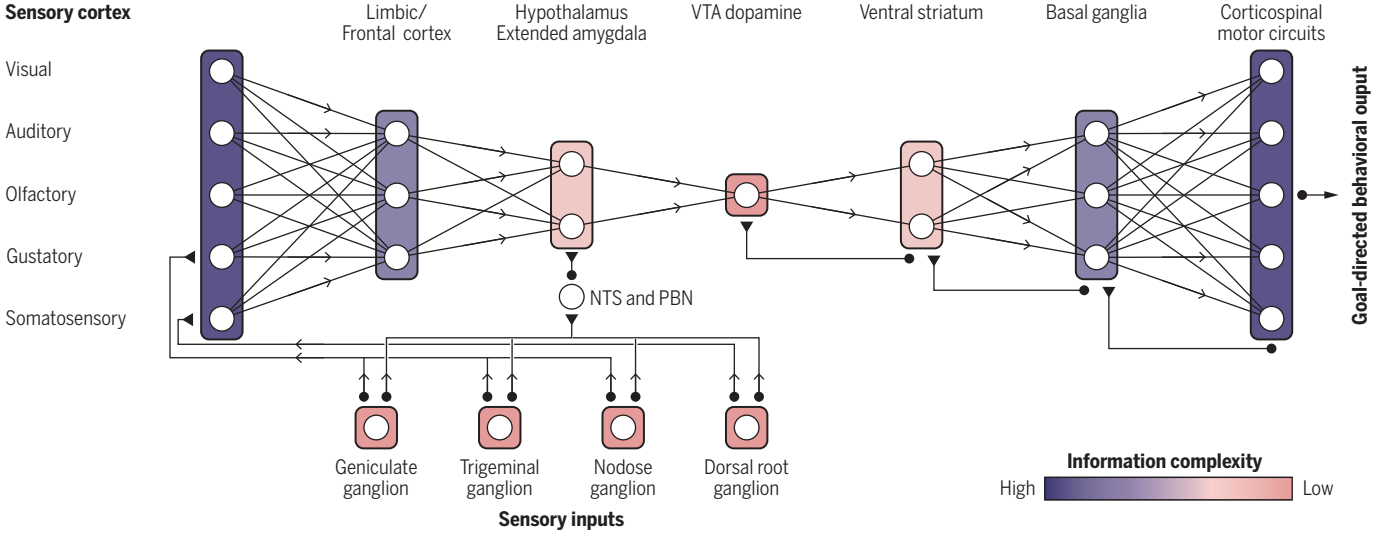
Hypothalamic and extended amygdala circuits are positioned upstream of the VTA dopamine system

The LHA and extended amygdala glutamatergic and GABAergic cells also project directly to the VTA and predominantly synapse onto VTA GABAergic neurons (*32*, *42*) to directly bidirectionally control their activity, indirectly regulate dopamine release (*43*), and bidirectionally control motivated behavioral states (*42*). Additionally, optogenetic stimulation of LHA inhibitory inputs (*44*) or LHA GABAergic neurons (*45*) produces robust self-stimulation behavior but can also generate feeding-like motor patterns when noncontingent optogenetic stimulation is used (Fig. 3). These distinct motivation-associated phenotypes may be driven by specific cell ensembles or by molecularly defined LHA cell types (*46*–*48*). Consistent with this, largely nonoverlapping populations of LHA GABAergic neurons display increases in their activity time-locked to either reward approach or reward consumption (*45*) as well as cue-induced food seeking. These and other findings (*49*) suggest an important role of these cells also in coding of learned associations. Bulk optogenetic stimulation of LHA projections to the VTA produces self-stimulation behavior, whereas stimulations of fibers that traverse more posterior to the peri-Locus Coeruleus and neighboring brain areas produce motor actions associated with feeding (*50*), further suggesting an important role of LHA-VTA circuits in controlling motivation.

Glutamatergic neurons with the LHA seemingly produce opposing effects on reward- and feeding-related behaviors. Optogenetic stimulation of these cells immediately disrupts consumption of fluids and food in a frequency-dependent fashion (*46*). Ablation of these neurons also promotes excessive feeding and body weight gain, and optogenetic inhibition of LHA glutamatergic terminals within the LHb increases licking and drinking behavior (*36*). LHA glutamatergic neurons are molecularly heterogeneous and are also composed of distinct cell types. For example, LHA glutamatergic neurons that project to the VTA are enriched in expression of *preprodynorphin*, whereas those that project to the LHb express other distinct marker genes (*38*). Although LHA glutamatergic and GABAergic neurons produce opposing effects on feeding and reward seeking, the activity dynamics of LHA glutamatergic neurons is qualitatively like LHA GABAergic neuron in some conditions. For example, LHA glutamatergic neurons also respond by showing increases in their activity during reward consumption (*38*, *46*), which is somewhat at odds with the effects of stimulating these neurons. Food restriction also increases whereas chronic access to highly palatable food reduces the activity of LHA glutamatergic neurons (*46*). Furthermore, circulating hormones associated with feeding and fasting such as leptin and ghrelin can also modulate the activity of LHA glutamatergic neurons in a projection-specific fashion (*38*). Collectively, LHA cell types serve as critical input to canonical brain reward circuits while also projecting to brainstem locomotor regions to likely coincide dopamine release with brainstem motor patterns needed for feeding, such as mastication (Fig. 3).

Homeostatic signaling and circuits to modulate motivation

To transform physiological needs for food or water into motivation signals required to execute feeding and drinking behavior, neurons that directly respond to hunger or thirst interface with canonical brain reward circuits such as the VTA, extended amygdala, and LHA. Agouti-related protein (AgRP)–expressing neurons in



CREDIT: A. FISHER/*SCIENCE* BASED ON G. K. STUBER

Fig. 1. Transforming sensory input into motivated actions. Sensory input arrives through genetically hardwired circuits that convey distinct sensory qualities directly to the hypothalamus and hindbrain, as well as through indirect projections to sensory cortices. Within sensory cortices, important features are extracted and conveyed to neural networks in limbic and frontal cortices responsible for aggregating sensory modalities and learning associations between distinct sensory features such as the hippocampus and prefrontal cortex. Within these structures, causal associations between sensory features, location, and biological needs are established and conveyed back to subcortical systems critical for innate behaviors such as feeding and drinking. These innate behavioral circuits can tap into the VTA dopamine system to reinforce ongoing actions and associations in which reward-prediction signals are established by the aggregate VTA inputs. Dopamine release in the ventral striatum can then coincide with ongoing activity in the ventral striatum and eventually throughout the basal ganglia regions and motor cortex, which are critical for the execution of goal-directed actions and movements.

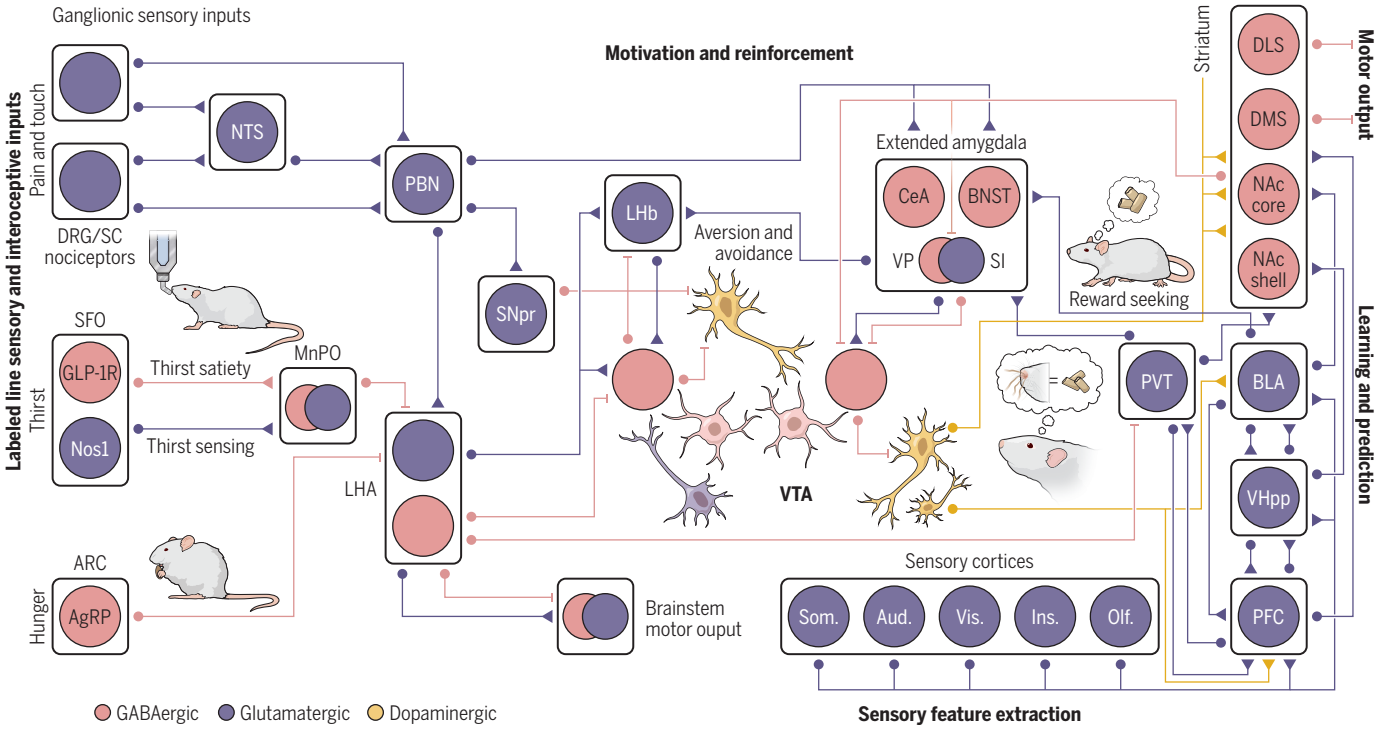


Fig. 2. Circuit connectivity linking circuits established for feeding, drinking, and touch to brain reward circuitry. Genetically defined neurons in areas such as the NTS, SFO, and arcuate hypothalamus (ARC) receive information related to ongoing sensory input from the skin and interoceptive systems that innervate organ systems. NTS, SFO, and ARC neurons indirectly project to the LHA's and related regions' excitatory and inhibitory neurons. Some LHA neurons project directly to brainstem motor circuits to drive reflexive behaviors, whereas some neurons project directly to the VTA. Excitatory LHA neurons project heavily to the LHb, which in turns negatively regulates the VTA dopamine system. LHA inhibitory neurons project directly to VTA, where they disinhibit the VTA dopamine system to enhance motivation for ongoing needs. DRG/SC, dorsal root ganglia/spinal cord; CeA, central amygdala; BNST, bed nucleus of the stria terminalis; VP/SI, ventral pallidum/substantia innominata; DLS, dorsolateral striatum; DMS, dorsomedial striatum; GLP-1R, glucagon-like peptide 1 receptor; PVN, paraventricular nucleus of the thalamus; VHipp, ventral hippocampus; Som, somatosensory cortex; Aud, auditory; Vis, visual; Ins, insular; Olf, olfactory.

the arcuate nucleus of the ventral hypothalamus serve as a critical rheostat for hunger states. AgRP neurons show increased activity under food deprivation, which decreases when mice encounter food (*51*, *52*). Activation of AgRP neurons produces feeding (*53*, *54*) and promotes conditioned place aversion (*51*), which is consistent with the role of AgRP neurons in transmitting hunger and potentially other aversive states. AgRP neuron activity also decreases in response to intragastric caloric delivery (*55*), suggesting that they are a critical node for transmitting interoceptive signals from the gut to the brain (*56*). The AgRP projections broadly target many subcortical systems; however, individual AgRP neurons tend to project to individual brain regions, such as the extended amygdala and LHA (*57*). Disrupting caloric sensing by AgRP neurons, and thus blunting their activation by food, results in reductions in feeding-associated dopamine release in the NAc (*58*). AgRP activation also potentiates VTA dopaminergic activity (*59*) and release in the NAc (*60*) but does not increase dopamine in the absence of food. Thus, AgRP neurons likely relay physiological homeostatic signals related to hunger indirectly to the dopamine system to enhance or reduce motivated feeding behavior.

Thirst signals to drive motivation for drinking are relayed to the VTA through fluid osmolarity-sensing neurons in subfornical organ (SFO)–expressing *nitric oxide synthase 1* (*Nos1*) (*61*, *62*). SFO^{Nos1} neurons project throughout the lamina terminalis, including the median preoptic nucleus (MnPO), which in turn sends inhibitory inputs to the LHA. LHA GABAergic neurons that project to VTA are activated by intragastric or oral infusions of water, as is the release of dopamine in downstream target regions, including the NAc (*22*). Stimulating SFO neurons chemogenetically boosts VTA dopaminergic responses to water-related cues, whereas inhibiting SFO decreases VTA dopaminergic activity (*63*). Collectively, VTA dopamine neurons are polysynaptically downstream from thirst- and hunger-sensing neurons to permit incoming interoceptive sensory to regulate their activity needed for motivation through the hypothalamus.

Circuits for touch and taste to modulate motivation

In addition to motivated actions such as feeding and drinking that require approach and consumption to reduce homeostatic drives, neural systems that encode danger, pain, and taste aversion also regulate canonical brain reward systems to promote behavioral avoidance. Genetically identified parabrachial nucleus (PBN) (*64*) neurons in the brain stem receive direct input from the nucleus tractus solitarius (NTS), dorsal laminae of the spinal cord, and forebrain regions. The NTS and PBN receive input from vagal, geniculate ganglion, and trigeminal sensory ganglion neurons that innervate peripheral organs (vagus) or mouth and tongue (trigeminal and geniculate ganglia). The NTS and PBN also indirectly receive input from nociceptors (TrpV1-lineage neurons) in dorsal root ganglia through the spinal cord. A genetically defined population of PBN neurons that express the gene for calcitonin gene–related peptide (*Calca*) show robust increases in their activity in response to cutaneous and visceral stimuli associated with pain and illness (*65*). Excitatory PBN neurons project broadly to the midbrain and forebrain, including the extended amygdala and LHA. PBN neurons form monosynaptic inputs on LHA glutamatergic neurons to reduce feeding and promote behavioral avoidance of nociceptive stimuli (*66*) to presumably regulate the midbrain dopamine system through the $PBN^{Glutamate} \rightarrow LHA^{Glutamate} \rightarrow LHb^{Glutamate} \rightarrow VTA^{GABA} \rightarrow VTA^{Dopamine}$ circuit.

Additionally, PBN signals that encode nociceptive stimuli can also reach the VTA through feedforward inhibition from the substantia nigra pars reticulata (Snpr) ($PBN^{Glutamate} \rightarrow SnR^{GABA} \rightarrow VTA^{Dopamine}$) (*67*). Whereas aversive sensory stimuli can indirectly suppress VTA dopamine and promote avoidance, pleasant sensory stimuli, such as light touch (*68*), increase NAc dopamine release, suggesting that further delineation of cutaneous sensory-reward systems is needed.

Learned associations to drive motivation

Genetically specified cells and circuits within structures such as the hypothalamus and PBN receive sensory input and regulate the activity of VTA dopamine neurons to promote motivated behaviors. Although these systems are critical for producing stereotypical behavioral responses, their sheer simplicity cannot alone account for how learning and other cognitive processes such as decision-making influence brain reward circuitry. Cortical structures including the hippocampus, prefrontal cortices, and basolateral amygdala (BLA) play crucial roles in learning associations between environmental stimuli and desirable motivated outcomes (*69*–*72*). Cortical neuronal networks within these regions facilitate the writing, storage, and recall of learned information (*73*). They also execute specific instinctual programs through their excitatory projections to the striatum (*74*, *75*), extended amygdala (*76*), hypothalamus (*77*), and midbrain (*72*). Neural coding of environmental stimuli, repeatedly paired with valued motivational outcomes such as licking to receive a food or water reward, are broadly coded throughout the brain (*78*) but are enriched in regions such as the orbitofrontal and prelimbic cortex (*72*, *79*). These learning- and prediction-layer cortical areas receive direct information from sensory cortical areas, including the insular cortex (*80*–*83*). Insula neurons can individually code for distinct salient sensory features such as sweet or bitter tastes. These are conveyed in a topographical fashion to learning and prediction regions such as the amygdala (*81*). Insula neurons also provide direct input to the NTS to control motivational vigor (*84*) and provide feedback to NTS sensory neurons (*85*). Collectively, insula and likely other sensory cortices direct motivated behaviors by routing various sensory features to learning and prediction regions, such as prefrontal cortex and amygdala, to learn higher-order and multisensory representations related to specific drives, which in turn project to subcortical systems to execute motivation-specific behavioral patterns related to approach and avoidance.

Outlook

This neurocircuit framework that links environmental and interoceptive stimuli to genetically defined cell types to produce motivated behaviors puts forward new testable hypotheses and unresolved questions. First, this model suggests that highly specific sensory features with no associated valence should not be represented by VTA dopamine neurons or other areas immediately upstream. Highly specific sensory information (for example, an individual odor or taste) may promote activity in the dopamine system only if that quality of the sensory stimulus is associated over time with rewarding or aversive outcomes such as caloric satiation. Tastes with associated motivational value may be represented in areas such as the insula cortex, which then indirectly reaches the VTA in a highly distilled form; however, taste reactivity (innate orofacial movements associated with sweet or bitter tastes) does not necessarily require cortical circuitry (*86*). Second, neurocircuits for motivated behaviors such as feeding or drinking should dynamically regulate the VTA dopamine system when a particular biological need is high. This process likely requires ongoing circuit adaptation and plasticity to allow for activity in some systems to be prioritized when needs are high and deprioritized when they are low. Signaling systems such as circulating hormones, cytokines, and other neuromodulators also likely play an important role in regulating circuit-specific activity to meet ongoing physiological demands (*87*–*89*). Last, interoceptive signals, such as the activity of genetically defined vagal sensory neurons in the nodose ganglia, can drive learning and specific motivated behaviors, but how and where in the brain learned associations between interoceptive and external cues are integrated is likely specific for distinct types of motivated actions and requires more in-depth study.

Although some genetically defined cell types, especially within the hypothalamic regions, are well described, much work is still needed to

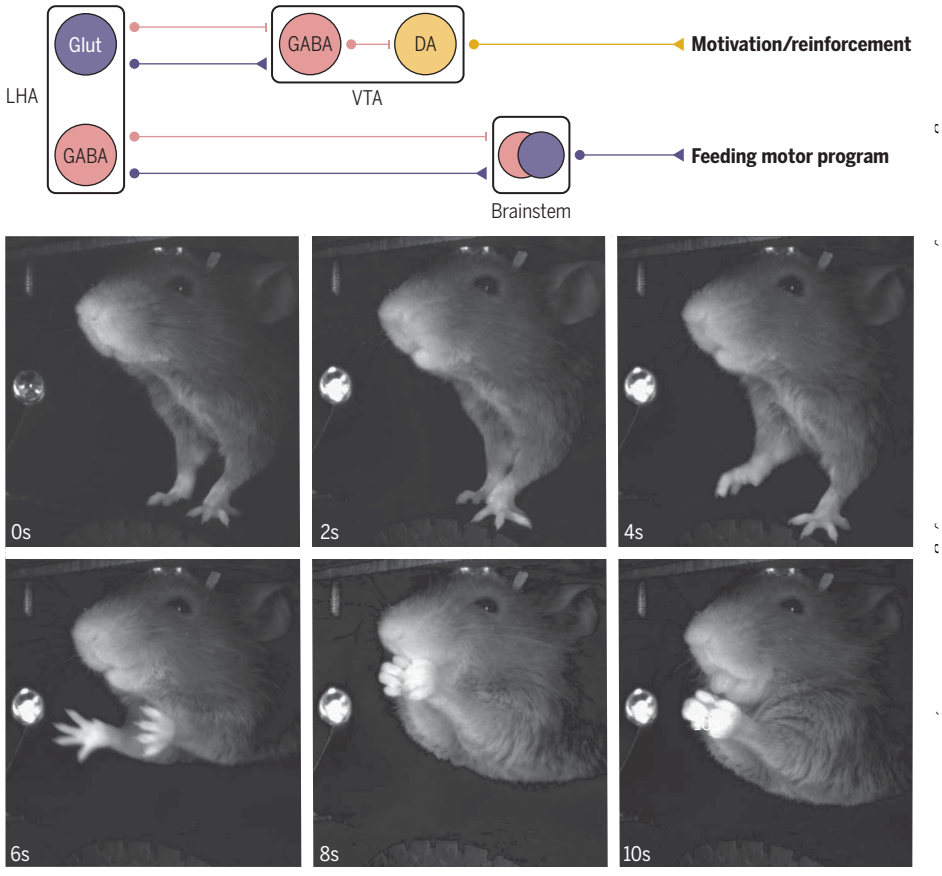


Fig. 3. Optogenetic stimulation of LHA GABAergic neurons is reinforcing but also produces complex motor patterns associated with feeding. Mice will readily perform an arbitrary movement, such as a nose poke or lever press, to receive optical stimulation of LHA GABAergic neurons. The reinforcing effects of LHA GABAergic neuron activity is largely through these cells' inhibitory drive onto VTA GABAergic neurons. Inhibiting VTA GABA neurons increases the activity of neighboring dopamine neurons to reinforce ongoing action. Meanwhile, LHA GABAergic projections also directly innervate the brainstem motor regions to activate innate feeding motor programs. The video still images show a head-fixed mouse starting to receive optogenetic stimulation of LHA GABAergic neurons at 2 s. Upon the start of stimulation, mice first start showing oral facial movements, followed by complex motor behaviors, including bringing paws to mouth, mastication, and licking. When the stimulation is turned off, these motor behaviors immediately cease. DA, dopamine.

reconcile how genetically defined cell types form specified circuits to modulate motivation. This has been a slow process because typically, cell types have been defined by using only one to a few highly expressed genes coupled with their anatomical connectivity. More comprehensive approaches, such as single-cell RNA-sequencing or spatial transcriptomics used in conjunction with monitoring cell-specific neural dynamics during motivated behavior, will lead to a comprehensive understanding of molecular features of cell type coupled with their intrinsic function (*90*). In addition, although optogenetic and chemogenetic studies have been instrumental in determining how genetically defined cell types can initiate or reduce ongoing motivated behaviors such as feeding, these techniques are not yet capable of recapitulating complex activity dynamic patterns of large-scale cell ensembles. Temporally patterned optogenetic stimulation of individual neurons to reproduce naturally occurring ensemble dynamics promises to provide detailed mechanistic information by which select subsets of motivationally relevant neurons and their patterned activity translate into motivational action. Collectively, the understanding of brain circuit mechanisms for motivation marches forward. Continued study and synthesis of information relating cell types and circuits to precise aspects of motivated behavior will not only deepen our understanding of the neurobiology of behavior but also pave the way for new interventions for a wide range of motivation-related disorders.

REFERENCES AND NOTES

1. J. R. Stellar, E. Stellar, *The Neurobiology of Motivation and Reward* (Island Press, 1985).
2. I. P. Pavlov, *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (Oxford Univ. Press, 1927).
3. C. B. Ferster, B. F. Skinner, *Schedules of Reinforcement* (Appleton-Century-Crofts, 1957).
4. J. Olds, *Science* **127**, 315–324 (1958).
5. E. Stellar, *Psychol. Rev.* **61**, 5–22 (1954).
6. C. Lüscher, P. H. Janak, *Annu. Rev. Neurosci.* **44**, 173–195 (2021).
7. J. A. McHenry *et al*., *Nat. Neurosci.* **20**, 449–458 (2017).
8. L. Yin *et al*., *Neuron* **110**, 3000–3017.e8 (2022).
9. S. Inoue *et al*., *Cell* **179**, 1393–1408.e16 (2019).
10. H. Lee *et al*., *Nature* **509**, 627–632 (2014).
11. D. Lin *et al*., *Nature* **470**, 221–226 (2011).
12. S. Stagkourakis *et al*., *Cell* **182**, 960–975.e15 (2020).
13. Y.-Y. Fang, T. Yamaguchi, S. C. Song, N. X. Tritsch, D. Lin, *Neuron* **98**, 192–207.e10 (2018).
14. D. C. Castro *et al*., *Nature* **598**, 646–651 (2021).
15. R. A. Wise, *Nat. Rev. Neurosci.* **5**, 483–494 (2004).
16. I. B. Witten *et al*., *Neuron* **72**, 721–733 (2011).
17. K. C. Berridge, *Psychopharmacology* **191**, 391–431 (2007).
18. W. Schultz, *J. Neurophysiol.* **80**, 1–27 (1998).
19. J. Y. Cohen, S. Haesler, L. Vong, B. B. Lowell, N. Uchida, *Nature* **482**, 85–88 (2012).
20. H. Jeong *et al*., *Science* **378**, eabq6740 (2022).
21. B. Engelhard *et al*., *Nature* **570**, 509–513 (2019).
22. J. C. R. Grove *et al*., *Nature* **608**, 374–380 (2022).
23. H.-C. Tsai *et al*., *Science* **324**, 1080–1084 (2009).
24. S. Lammel *et al*., *Nature* **491**, 212–217 (2012).
25. J. E. Markowitz *et al*., *Nature* **614**, 108–117 (2023).
26. J. W. de Jong *et al*., *Neuron* **101**, 133–151.e7 (2019).
27. C. M. Vander Weele *et al*., *Nature* **563**, 397–401 (2018).
28. R. A. Phillips III *et al*., *Cell Rep.* **39**, 110616 (2022).
29. R. van Zessen, J. L. Phillips, E. A. Budygin, G. D. Stuber, *Neuron* **73**, 1184–1194 (2012).
30. K. R. Tan *et al*., *Neuron* **73**, 1173–1183 (2012).
31. N. Eshel *et al*., *Nature* **525**, 243–246 (2015).
32. K. T. Beier *et al*., *Cell* **162**, 622–634 (2015).
33. M. Watabe-Uchida, L. Zhu, S. K. Ogawa, A. Vamanrao, N. Uchida, *Neuron* **74**, 858–873 (2012).
34. J. Tian *et al*., *Neuron* **91**, 1374–1389 (2016).
35. A. M. Stamatakis, G. D. Stuber, *Nat. Neurosci.* **15**, 1105–1107 (2012).
36. A. M. Stamatakis *et al*., *J. Neurosci.* **36**, 302–311 (2016).
37. I. Lazaridis *et al*., *Mol. Psychiatry* **24**, 1351–1368 (2019).
38. M. A. Rossi *et al*., *Neuron* **109**, 3823–3837.e6 (2021).
39. J. Tooley *et al*., *Biol. Psychiatry* **83**, 1012–1023 (2018).
40. S. Lecca *et al*., *eLife* **6**, e30697 (2017).
41. D. J. Barker *et al*., *Cell Rep.* **21**, 1757–1769 (2017).
42. J. H. Jennings *et al*., *Nature* **496**, 224–228 (2013).
43. E. H. Nieh *et al*., *Neuron* **90**, 1286–1298 (2016).
44. J. H. Jennings, G. Rizzi, A. M. Stamatakis, R. L. Ung, G. D. Stuber, *Science* **341**, 1517–1521 (2013).
45. J. H. Jennings *et al*., *Cell* **160**, 516–527 (2015).
46. M. A. Rossi *et al*., *Science* **364**, 1271–1274 (2019).
47. L. E. Mickelsen *et al*., *Nat. Neurosci.* **22**, 642–656 (2019).
48. Y. Wang *et al*., *Cell* **184**, 6361–6377.e24 (2021).
49. M. J. Sharpe *et al*., *Curr. Biol.* **27**, 2089–2100.e5 (2017).
50. R. A. M. Marino *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 8611–8615 (2020).
51. J. N. Betley *et al*., *Nature* **521**, 180–185 (2015).
52. Y. Chen, Y.-C. Lin, T.-W. Kuo, Z. A. Knight, *Cell* **160**, 829–841 (2015).
53. Y. Aponte, D. Atasoy, S. M. Sternson, *Nat. Neurosci.* **14**, 351–355 (2011).
54. M. J. Krashes *et al*., *J. Clin. Invest.* **121**, 1424–1428 (2011).
55. L. R. Beutler *et al*., *Neuron* **96**, 461–475.e5 (2017).
56. L. Bai *et al*., *Cell* **179**, 1129–1143.e23 (2019).
57. J. N. Betley, Z. F. H. Cao, K. D. Ritola, S. M. Sternson, *Cell* **155**, 1337–1350 (2013).
58. A. Reichenbach *et al*., *eLife* **11**, e72668 (2022).
59. C. M. Mazzone *et al*., *Nat. Neurosci.* **23**, 1253–1266 (2020).
60. A. L. Alhadeff *et al*., *Neuron* **103**, 891–908.e6 (2019).
61. C. A. Zimmerman *et al*., *Nature* **568**, 98–102 (2019).
62. V. Augustine *et al*., *Nature* **555**, 204–209 (2018).
63. T. M. Hsu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30744–30754 (2020).
64. J. L. Pauli *et al*., *eLife* **11**, e81868 (2022).
65. C. A. Campos, A. J. Bowen, C. W. Roman, R. D. Palmiter, *Nature* **555**, 617–622 (2018).
66. S. C. Phua *et al*., *Sci. Adv.* **7**, eabe4323 (2021).
67. H. Yang *et al*., *Nat. Neurosci.* **24**, 1402–1413 (2021).
68. L. J. Elias *et al*., *Cell* **186**, 577–590.e16 (2023).
69. A. Beyeler *et al*., *Neuron* **90**, 348–361 (2016).
70. J. S. Biane *et al*., *Nat. Neurosci.* **26**, 798–809 (2023).
71. J. M. Otis *et al*., *Nature* **543**, 103–107 (2017).
72. V. M. K. Namboodiri *et al*., *Nat. Neurosci.* **22**, 1110–1121 (2019).
73. V. M. K. Namboodiri, G. D. Stuber, *Neuron* **109**, 3552–3575 (2021).
74. J. Cox, I. B. Witten, *Nat. Rev. Neurosci.* **20**, 482–494 (2019).
75. X. Zhang *et al*., *Nat. Neurosci.* **24**, 1586–1600 (2021).
76. T. Yang *et al*., *Nature* **616**, 510–519 (2023).
77. N. Padilla-Coreano *et al*., *Nature* **603**, 667–671 (2022).
78. J. E. Allen, T. E. Sutherland, *Science* **364**, 738–739 (2019).
79. D. J. Ottenheimer, M. M. Hjort, A. J. Bowen, N. A. Steinmetz, G. D. Stuber, *eLife* **12**, RP84604 (2023).
80. V. M. K Namboodiri *et al*., *Curr. Biol.* **31**, 5176–5191.e5 (2021).
81. L. Wang *et al*., *Nature* **558**, 127–131 (2018).
82. Y. Livneh *et al*., *Nature* **546**, 611–616 (2017).
83. Y. Livneh, M. L. Andermann, *Neuron* **109**, 3576–3593 (2021).
84. H. Deng *et al*., *Cell* **184**, 6344–6360.e18 (2021).
85. H. Jin, Z. H. Fishman, M. Ye, L. Wang, C. S. Zuker, *Cell* **184**, 257–271.e16 (2021).
86. H. J. Grill, R. Norgren, *Brain Res.* **143**, 281–297 (1978).
87. M. E. Soden, J. X. Yee, L. S. Zweifel, *Nature* **619**, 332–337 (2023).
88. S. X. Zhang *et al*., *Nature* **597**, 245–249 (2021).
89. D. C. Castro *et al*., *Nature* **598**, 646–651 (2021).
90. S. Xu *et al*., *Science* **370**, eabb2494 (2020).

ACKNOWLEDGMENTS

I acknowledge Stuber Lab members for helpful discussion and feedback on the manuscript, M. Trzeciak and A. Gordon-Fennell for assistance with figure preparation, and the National Institute on Drug Abuse for support (DA032750, DA038168, and DA054317).
Competing interests: None declared. **License information:**
Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 17 July 2023; accepted 22 September 2023
10.1126/science.adh8287

REVIEW

Hypothalamic control of innate social behaviors

Long Mei[1], Takuya Osakada[1], Dayu Lin[1,2,3,4]*

Sexual, parental, and aggressive behaviors are central to the reproductive success of individuals and species survival and thus are supported by hardwired neural circuits. The reproductive behavior control column (RBCC), which comprises the medial preoptic nucleus (MPN), the ventrolateral part of the ventromedial hypothalamus (VMHvl), and the ventral premammillary nucleus (PMv), is essential for all social behaviors. The RBCC integrates diverse hormonal and metabolic cues and adjusts an animal's physical activity, hence the chance of social encounters. The RBCC further engages the mesolimbic dopamine system to maintain social interest and reinforces cues and actions that are time-locked with social behaviors. We propose that the RBCC and brainstem form a dual-control system for generating moment-to-moment social actions. This Review summarizes recent progress regarding the identities of RBCC cells and their pathways that drive different aspects of social behaviors.

The medial hypothalamus houses three heavily interconnected nuclei—the medial preoptic nucleus (MPN), the ventrolateral part of the ventromedial hypothalamus (VMHvl), and the ventral premammillary nucleus (PMv)—that constitute the reproductive behavior control column (RBCC) (*1*). The RBCC is indispensable for all social behaviors that subserve reproduction, including mating, parenting, and fighting. It is part of the larger social behavior network (SBN) that was initially proposed by S. Newman in 1999 on the basis of rodent studies (*2*) and was later extended by J. L. Goodson to other vertebrate species (*3*). Anatomically, the RBCC is highly evolutionarily conserved and recognizable in mammals, birds, reptiles, amphibians, and fish (*3*, *4*). Immediate early gene mapping suggests similar response patterns during social behaviors across species (*3*, *5*). In this Review, we summarize recent studies, mainly in rodents, regarding the molecular identity, functions, and responses of RBCC cells. RBCC function in nonrodent species has been reviewed in other work (*4*, *6*).

This Review will be separated into two parts. The first part summarizes recent studies regarding the functions and responses of molecularly defined RBCC cells during each social behavior. We will also offer our views about the general principles of RBCC organization and the relationship among different social circuits. The second part of the Review discusses how the RBCC controls each stage of social behaviors, from increasing social exploration and maintaining social interest to promoting consummatory social actions. In particular, we propose that the RBCC acts as a "permissive gate" to allow specific brainstem and/or spinal cord motor circuits to respond to immediate and simple sensory inputs to drive moment-to-moment social actions.

[1]Neuroscience Institute, New York University Langone Medical Center, New York, NY 10016, USA. [2]Department of Psychiatry, New York University Langone Medical Center, New York, NY 10016, USA. [3]Department of Neuroscience and Physiology, New York University Langone Medical Center, New York, NY 10016, USA. [4]Center for Neural Science, New York University, New York, NY 10016, USA.
*Corresponding author. Email: dayu.lin@nyulangone.org

The role of the RBCC in sexual, parental, and aggressive behaviors

Male sexual behaviors

Sexual behaviors differ qualitatively between the sexes. Thus, the male and female sexual behavior circuits have little in common. For males, the medial preoptic area (MPOA), which includes the MPN and its surrounding region, has been recognized as a site of paramount importance since the early 1960s (*7*, *8*). Recently, single-cell and nucleus RNA sequencing have revealed a high molecular heterogeneity of MPOA cells, leading to a series of studies to refine the molecular identities of mating-related MPOA subpopulations (*9*, *10*). In 2018, Wei *et al.* showed that optogenetic activation of MPOA *estrogen receptor alpha* (*Esr1*)–expressing cells ($MPOA^{Esr1}$) can elicit male mounting, albeit with relatively low efficiency (50% of animals) (*11*). In 2021, Karigo *et al.* reported that activating γ-aminobutyric acid–expressing (GABAergic) $MPOA^{Esr1}$ ($MPOA^{Esr1\cap VGAT}$) cells can induce male mounting with 100% efficiency (*12*). Most recently, Bayless *et al.* pinpointed *tachykinin receptor 1* (*Tacr1*)–expressing cells ($MPOA^{Tacr1}$), likely a subset of $MPOA^{Esr1\cap VGAT}$ cells (*10*), as the key population for male sexual behaviors (*13*). Optogenetic activation of $MPOA^{Tacr1}$ cells was able to induce intromission-like behaviors toward a toy mouse (*13*). Importantly, Tacr1 is itself functionally critical for the behavior because antagonizing MPOA Tacr1 suppressed male mounting (*13*).

The role of the VMHvl in male sexual behaviors appears to be relatively minor. Optogenetic activation of $VMHvl^{Esr1}$ cells at a low intensity induced mounting in male mice, but this behavior was not coupled with ultrasonic vocalization, which typically accompanies male-female interaction, and hence was interpreted as a dominance behavior instead of a sexual behavior (*12*, *14*). Although ablating VMHvl *progesterone receptor* (*PR*)–expressing cells ($VMHvl^{PR}$) or chemogenetically inhibiting $VMHvl^{Esr1}$ cells (*Esr1* and *PR* overlap nearly 100% at the VMHvl) suppressed male mounting (*12*, *15*), optogenetic inhibition of $VMHvl^{Esr1}$ cells did not disrupt ongoing copulation, suggesting that the role of the VMHvl in male sexual behavior is likely limited to the initiation phase (*14*). This functional result is supported by the response pattern of VMHvl cells: Whereas the cells increase activity during mounting, they decrease activity during intromission (*16*, *17*).

The PMv receives dense inputs from the medial amygdala, a social odor and pheromone processing region, and thus has been considered the sensory relay of the RBCC (*18*). Indeed, the PMv expresses abundant c-Fos after conspecific odor presentation and shows a moderate activity increase during the male-female investigation (*17*, *19*). However, PMv cells do not respond during any phases of male copulation, and PMv deficits fail to impair male reproduction (*17*, *20*, *21*).

Thus, the MPOA is likely the primary site that mediates male mating, whereas the VMHvl and PMv seem to play minor roles in the early phase of male-female interaction (Fig. 1). $MPOA^{Tacr1}$ cells represent the most refined population for male sexual behaviors (Fig. 2).

Female sexual behaviors

Female sexual behavior, in comparison, is notably simple in its motor output. As males mount, females stay stationary with their back arching downward to facilitate penile insertion, a posture known as lordosis (*22*). Since the 1970s, many lesion and stimulation experiments have demonstrated a critical role of the VMHvl in female sexual behaviors (*22*). Similar to studies in males, recent efforts focused on refining the molecular identity of VMHvl cells relevant to female sexual behaviors. Toward this goal, $VMHvl^{PR}$ cells were identified as necessary for female sexual receptivity in mice (*15*, *23*). However, activating $VMHvl^{PR}$ cells failed to facilitate lordosis (*23*). This negative result is likely due to the fact that the female VMHvl contains two subdivisions, which both express abundant *PR* and *Esr1*, but only the lateral subdivision (VMHvll) is strongly activated during female sexual behaviors (*24*). When the VMHvl cells that express *Cholecystokinin A receptor* (*Cckar*), a lateral subdivision–specific gene, are chemogenetically or optogenetically activated, female sexual receptivity rapidly increases (*10*, *25*). In mice, female sexual receptivity is tightly coupled with ovulation, that is, the estrus cycle. The in vivo responses and physiological properties of $VMHvll^{Cckar}$ cells, including intrinsic excitability and synaptic inputs, vary over the estrus cycle, suggesting that these cells probably support the cyclic changes in female sexual behaviors (*25*).

The PMv is also critical for female reproduction, but its role differs from that of the VMHvl. Specifically, the PMv enables metabolic signals to modulate the reproductive neuroendocrine axis. Leptin, an adipocyte-derived hormone, signals energy reserve levels and triggers sexual

maturation by acting on PMv leptin receptors (Lepr) (*26*). Upon sensing leptin, PMv^{Lepr} cells activate *kisspeptin*-expressing cells to release gonadotropin-releasing hormone and trigger the onset of puberty (*20*, *26*). In adults, PMv impairment decreases ovulation frequency, an abnormality that could occur naturally during leptin deficiency caused by food deprivation (*20*, *27*). Thus, PMv modulates sexual readiness based on the energy reserve level. During food shortages, the PMv signals the hypothalamic-pituitary-ovarian axis to slow down female reproduction.

The MPOA has been consistently found to suppress female receptivity, possibly through its strong inhibitory inputs to the VMHvl (*12*, *28*). MPOA lesion or site-specific *Esr1* knockdown increases lordosis, whereas electric stimulation of the MPOA has the opposite effect (*29*–*33*).

Thus, contrary to the male mating circuit, female sexual behaviors are mainly mediated by the VMHvl and PMv, whereas the MPOA likely plays a negative role (Fig. 1). VMHvll^{Cckar} and PMv^{Lepr} cells represent the most refined populations for female sexual behavior thus far (Fig. 2).

Parental behaviors

In rodents, all parental behaviors, except nursing, can be exhibited by both males and females with some quantitative differences (*34*). Thus, perhaps unsurprisingly, the hypothalamic regions that control parental behaviors are largely similar between the sexes. Since 1974, the MPOA has been recognized as a crucial region for controlling maternal behaviors (*35*, *36*). More recently, Wu *et al.* identified MPOA cells that express *galanin* (MPOA^{Gal}) as a critical population for regulating parental behaviors, especially grooming, in male and female mice (*37*). Subsequently, Wei *et al.* and our study collectively found that both male and female MPOA^{Esr1} cells in mice are necessary, sufficient, and naturally active during pup approach and retrieval (*11*, *38*). Later, multiplexed error robust fluorescence in situ hybridization (MERFISH) revealed that *calcitonin receptor* (*Calcr*), which is expressed in a subset of *Esr1*-positive cells, preferentially marks MPOA cells that are activated by parental behaviors (*9*). Following this finding, Yoshihara *et al.* reported that silencing MPOA^{Calcr} neurons or knocking down *Calcr* in the MPOA suppresses maternal behaviors (*39*). *Gal*, *Esr1*, and *Calcr* are expressed in both GABAergic and glutamatergic MPOA cells. However, the glutamatergic MPOA cells likely do not promote parental behaviors because optogenetic silencing of MPOA glutamatergic cells increases, not decreases, pup retrieval in female mice (*40*). Lastly, MPOA cells not only drive parental behaviors but also suppress hostile behaviors toward the pups. During lactation, MPOA^{Esr1} cell excitability increases, causing stronger inhibition of infanticide-driving cells in the principal nucleus of the bed nucleus of stria terminalis (BNSTpr) and enabling the drastic switch from hostile to caring behaviors toward the young (*41*).

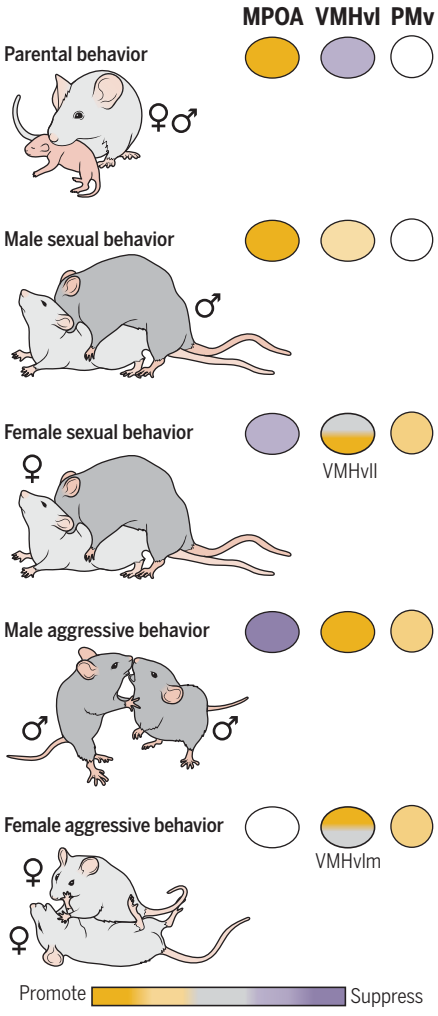


Fig. 1. The role of RBCC regions in various social behaviors. White indicates that the function of the region is to be determined. Gray indicates the region has no function. Purple indicates that the region suppresses the behavior. Yellow indicates that the region promotes the behavior. Note that the MPOA and the VMHvl and PMv generally play opposite roles in each social behavior.

In contrast to the pivotal role of the MPOA in parental care, the VMHvl and PMv appear to be dispensable. VMHvl inactivation or PMv lesion did not cause any deficit in maternal care (*24*, *42*). In vivo recordings found no activity changes of VMHvl^{Esr1} cells during maternal behaviors (*24*). Recent studies further suggest a negative role of the VMHvl in parental behaviors. Activating VMHvl cells that are connected with the MPOA suppressed pup investigation in virgin female mice (*41*). Similarly, when the projection from *urocortin-3* (*Ucn3*)–expressing cells in the perifornical region (PeFA^{Ucn3}) to the VMH was activated, the female mice spent less time investigating pups (*43*).

Altogether, present data support the idea that the MPOA, but not the VMHvl and PMv, is the key RBCC region that drives parental behaviors in both sexes (Fig. 1). MPOA^{Calcr} cells represent the most refined population for parental behaviors (Fig. 2).

Aggressive behaviors

Although the tendency to attack differs between the sexes, its motor pattern, at least in rodents, seems to be similar (*44*). Thus, the neural substrates of aggression are qualitatively similar in males and females. Studies from the past decade firmly established the central role of the VMHvl in male and female aggression (*45*). Inactivating or ablating VMHvl cells abolishes natural intermale aggression and maternal aggression in mice (*14*–*16*, *24*). Conversely, activating VMHvl cells, especially those expressing *Esr1* or *PR*, promotes attack toward both natural and suboptimal targets (*14*, *16*, *24*), regardless of the subject's social status, housing condition, or testing context (*46*). Simultaneous recording of 13 limbic regions, including five in the hypothalamus, revealed the VMHvl as the region with the largest and fastest activity increase during attack onset, highlighting its crucial role in the behavior (*17*). Furthermore, recent works revealed the flexibility of VMHvl cell responses. For example, with sexual experience, the male- and female-induced activation patterns in the VMHvl become more distinct (*47*). With winning experience, the VMHvl cells show long-term potentiation of excitatory synaptic inputs (*48*). When an arbitrary motor action, for example, a nose poke, is associated with future opportunities to attack, VMHvl cells increase activity before poking (*49*). Lastly, Yang *et al.* recently reported increased VMHvl cell activity when animals witness fights between others (*50*). Thus, the VMHvl cells carry diverse aggression-related information, including aggressive motivation, aggression-provoking sensory cues, the motor execution of attacks, and one's own fighting experiences and those of others. Notably, only the posterior VMHvl is related to aggression; the anterior VMHvl mediates conspecific self-defense and social fear (*51*, *52*).

In female mice, only the medial VMHvl (VMHvlm) is relevant for aggression (*24*). Using activity-dependent single-cell RNA sequencing (Act-seq), Liu *et al.* identified *neuropeptide Y receptor Y2* (*Npy2r*), a VMHvlm-biased gene, as a genetic marker for the female aggression population (*53*). When *Npy2r*$^{+}$, but not *Npy2r*$^{-}$ *Esr1*$^{+}$, VMHvl cells were optogenetically activated, virgin female mice attacked various social targets, including adult males (*53*). VMHvl^{Npy2r} cells show reproductive state–dependent activity changes, with the highest response occurring during lactation, when the level of female aggression peaks (*53*).

The PMv is also a critical site for both male and female aggression. Several recent studies targeted *dopamine transporter* (*DAT*)–expressing cells in the PMv (PMv^{DAT}) and showed that

aggression can be bidirectionally modulated in male and female mice (*21*, *54*, *55*). In vivo recordings show that PMv cells, like VMHvl cells, respond preferentially to aggression-provoking social cues (*17*, *21*, *56*). However, the PMv differs from the VMHvl in that it responds more during social investigation than attack (*17*, *21*). Lesioning the PMv reduces aggression-induced c-Fos in the VMHvl substantially, suggesting that the PMv is likely to be upstream of the VMHvl, relaying conspecific cues (*42*). Of note, regardless of sex, PMv^{DAT} cells are quiescent and hyperpolarized in non-aggressive mice and become spontaneously active and depolarized in aggressive mice (*54*, *55*), suggesting that the aggressiveness of an animal could be encoded by the biophysical properties of PMv cells. It is worth mentioning that although *DAT* is a good marker for PMv because it is distinctively expressed in the PMv among neighboring regions, PMv^{DAT} cells do not appear to synthesize dopamine, and it remains unclear whether *DAT*$^{+}$ cells are more involved in aggression than *DAT*$^{-}$ cells in the PMv (*56*).

Recent studies support an aggression-suppressing role of MPOA^{Esr1} cells through their inhibition of VMHvl glutamatergic cells (*12*, *28*). Rostral MPOA^{Esr1} cells in males are activated by females, whereas caudal MPOA^{Esr1} cells preferentially respond to superior male opponents (*28*). Thus, the MPOA could suppress aggression toward inappropriate targets in various contexts. Indeed, when MPOA^{Esr1}-to-VMHvl terminals are optogenetically inactivated, male mice attack social targets indiscriminately, including superior males, and end up inflicting more defeat (*28*).

Present data suggest that the VMHvl and PMv are central for male and female aggression, whereas the MPOA suppresses this behavior (Fig. 1). In terms of cell populations, VMHvl^{Esr1} cells in males, which are equivalent to VMHvl^{PR} cells, and VMHvl^{Npy2r} cells in females are, at present, the best-defined populations for aggression (Fig. 2).

General organization of RBCC cells

Several general conclusions emerge when considering the circuits for all social behaviors together. First, the VMHvl and PMv have similar social functions that are opposite to those of the MPOA (Fig. 1). For example, the VMHvl and PMv promote aggression in both sexes and sexual behaviors in females, whereas MPOA activation suppresses these behaviors (*12*, *14*–*16*, *24*, *25*, *28*, *53*). The MPOA promotes parental behaviors, whereas the VMHvl reduces them (*11*, *37*–*39*, *41*, *57*). The only exception is male sexual behavior, the initiation of which is promoted by both the MPOA and VMHvl, although the MPOA increases, whereas the VMHvl decreases, activity during intromission (*12*, *14*, *17*). The functional relationship between the MPOA, PMv, and VMHvl is probably rooted in the neurotransmitter expressed by the cells. Cells in the PMv and VMHvl are overwhelmingly glutamatergic, whereas socially relevant cells in the MPOA are likely GABAergic (*12*, *40*). Because these three regions are heavily reciprocally connected, activating MPOA social cells should suppress PMv and VMHvl cells, whereas PMv and VMHvl cells should facilitate activation of each other (*28*, *56*).

Second, cells supporting distinct social behaviors often occupy different subregions within the VMHvl or MPOA (Fig. 2). For example, the lateral and medial VMHvl mediate female sexual and aggressive behaviors, respectively (*24*, *25*, *53*). Rostral and caudal MPOA cells in males are activated during female and dominant male interactions, respectively (*28*). *Fos* catFISH (compartment analysis of temporal activity by fluorescence in situ hybridization) suggests that mating- and parental care–induced *c-Fos* expression is largely distinct in the MPOA, and the former could be more medially located [see figure 2 in (*37*)]. At the single-cell level, most cells show biased, though not necessarily exclusive, responses during one social behavior (*16*, *38*, *47*, *58*). Thus, individual MPOA and VMHvl cells are likely preferentially involved in one social behavior, and cells with similar response patterns tend to cluster. Whether such topographic organization of social behaviors also exists in the PMv remains to be elucidated.

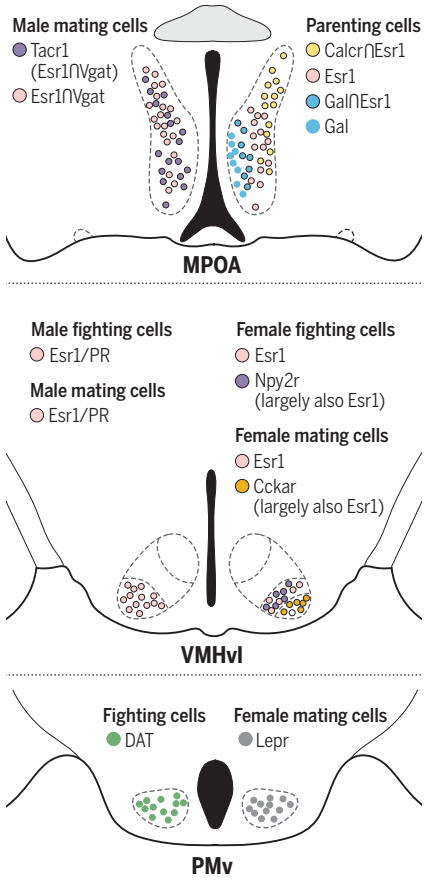


Fig. 2. The known molecular markers for social behavior–relevant cells in the RBCC.

Third, fighting- and mating-related cells in the RBCC antagonize each other in both sexes. In males, activating mating-related MPOA^{Esr1} cells suppresses intermale aggression, whereas activating fighting-related VMHvl^{Esr1} cells suppresses male-female mounting and female urine-elicited ultrasonic vocalization (*12*, *28*). In females, activating fighting-related VMHvl^{Npy2r} cells suppresses sexual receptivity (because the female attacks the male), whereas activating mating-related VMHvll^{Cckar} cells suppresses female aggression (*25*, *53*). This relationship is not simply due to motor incompatibility. In lactating females, VMHvll^{Cckar} cell activation could not increase female sexual receptivity, possibly because of changes in the downstream circuit, but remained effective in suppressing maternal aggression (*25*).

The relationship between parenting-related cells and mating- and fighting-related cells is less understood because these three behaviors are rarely examined together. Limited evidence suggests that these cells could operate independently. For example, after VMHvll^{Cckar} cells are inactivated, there is no change in maternal behaviors (*25*). Rather than counteracting with mating- and fighting-related cells, parenting-related cells form strong mutual inhibition with infanticide-related cells: Activating infanticide-related cells suppresses parenting-related cells and vice versa (*41*). These circuit relationships make sense when considering the social target of each behavior. Because mating and fighting are both directed to adult conspecifics with similar physical features, the mutual inhibition between these circuits ensures one behavior output dominates. By contrast, parenting behaviors and mating and fighting behaviors are activated by highly distinct social targets, making cross-activation unlikely and, hence, mutual inhibition unnecessary.

Lastly, as the number of studies increases, the "molecularly defined populations" for each behavior will keep growing. *Gal*-, *Esr1*-, and *Calcr*-expressing cells in the MPOA have all been found to be essential for driving parental behaviors on the basis of recording and functional studies (*37*–*39*) (Fig. 2). These cells are neither distinct nor identical. MPOA^{Gal} cells partially overlap with MPOA^{Esr1} cells but largely differ from MPOA^{Calcr} cells, whereas *Calcr* is expressed in a subset of MPOA^{Esr1} cells (*9*). How is it determined whether a gene truly marks the social behavior population or a random population containing some social behavior–relevant cells just by chance? Indeed, if a behavior is the dominant output of a brain region, activating a random subset of cells in the region, marked by any of thousands of expressed genes, may drive the behavior. We think two criteria should be met for a gene to be considered a relevant maker. First, the cells expressing the gene should show a higher activity change during the behavior than the gene-negative cells. Second, functional manipulation of the gene-positive cells should influence the behavior more

K. HOLOSKI/*SCIENCE* BASED ON MEI *ET AL.*

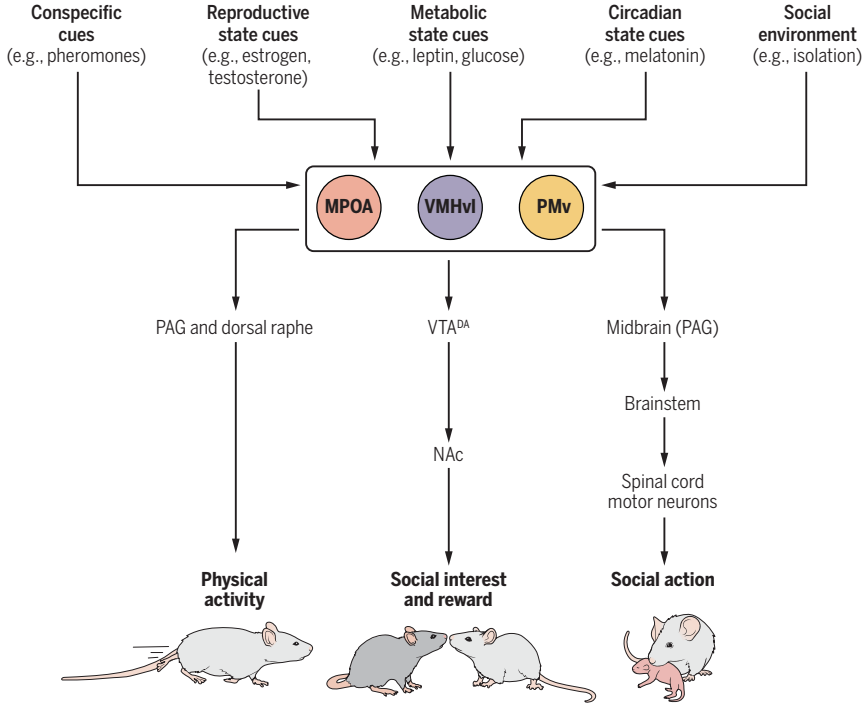


Fig. 3. Pathways extended from the RBCC that mediate various aspects of social behaviors.

profoundly than the cells that lack the expression. For example, *Esr1*-positive MPOA cells show higher responses than *Esr1*-negative cells during parental care (*38*). Activating *Esr1*-positive, but not *Esr1*-negative, cells in the VMHvl can drive male aggression (*14*). Additionally, it is unlikely, although not impossible, that a gene is distinctively expressed in the social behavior population coincidently. More plausibly, it is expressed for a reason, for example, to modulate the behavior output. Thus, a good gene marker is probably also functionally important for the behavior. Indeed, *Esr1* or *Calcr* knockout in the MPOA impairs parental behaviors, whereas antagonizing MPOA Tacr1 impairs male mating (*13*, *39*, *59*). Based on these rationales, receptors of neuropeptides or hormones that specifically modulate a social behavior could be the top candidates for marking specific social behavior populations.

The RBCC coordinates multiple aspects of social behaviors

We propose that the RBCC achieves social behavior control by (i) promoting exploration to increase the probability of social encounters, (ii) maintaining social interest when a social target is encountered, and (iii) permitting consummatory social actions at the appropriate times. In this section, we review the neural pathways downstream of the RBCC that mediate each aspect of these behavioral controls.

Social encounter probability

Animals must physically encounter each other to engage in social behaviors. We consider the first role of the RBCC to be to modulate an animal's physical activity, hence the likelihood of social encounters, based on an animal's internal state. In support of this idea, previous work has shown that animals increase locomotion when GABAergic MPOA cells (MPOA^{VGAT}), VMHvl^{Esr1} cells, or *NK2 homeobox 1* (*Nkx2-1*)–expressing VMHvl cells (VMHvl$^{Nkx2\text{-}1}$) are artificially activated (*60*–*63*). Conversely, loss of *Esr1* or *Nkx2-1* in the VMHvl reduces physical activity and causes obesity in female mice (*61*, *64*). VMHvl^{Esr1} projection to the dorsal raphe and MPOA^{VGAT} projection to the periaqueductal gray (PAG) appear important for RBCC control of physical activity (*62*, *63*) (Fig. 3).

The RBCC expresses abundant neuromodulator, neuropeptide, and hormone receptors, which allows various small molecules that are indicative of an animal's reproductive, metabolic, circadian, and other internal states to adjust the ongoing cell activity and, consequently, an animal's physical activity (Fig. 3). For the reproductive state, estrogen is a crucial hormone that signals ovarian activity. Estrogen supplementation substantially increases the excitability of MPOA and VMHvl cells (*65*, *66*). This estrogenic modulation likely underlies the increased spontaneous activity of VMHvl cells during estrus (*25*, *67*) and the increased activity of MPOA cells during motherhood (*38*, *41*). In addition to sex hormones, metabolic cues also modulate VMHvl^{Esr1} cell activity. For example, 24-hour fasting decreases the excitability of VMHvl^{Esr1} cells, which can be reversed with refeeding (*63*). Glucose may mediate these changes, because nearly all VMHvl cells are responsive to glucose in vitro (*68*). The circadian clock could also modulate the ongoing VMHvl activity. *Melanocortin receptor 4* (*MCR4*), a female VMHvl enriched gene, increases expression when estrogen levels are high, enabling nighttime melatonin to increase VMHvl cell activity and promote exploration during proestrus when the female is getting ready to mate (*69*, *70*). Beyond the internal cues, the social environment could also influence the ongoing activity of RBCC cells. Recently, Fukumitsu *et al.* found that prolonged social isolation decreases amylin and its receptor expression in the MPOA and reduces social-seeking behaviors (*71*). Thus, the ongoing RBCC cell activity is dynamically modulated by the internal state and external environment, which in turn adjusts an animal's physical activity level and the probability of social encounters.

Social interest level

Once the animals encounter each other, the next step is to recognize the identity of the social target. Although VMHvl and MPOA lesion or inhibition consistently change social preference, this likely does not reflect a deficit in sex discrimination, given that the manipulated animals often show reversed preference instead of no preference, which suggests that they remain capable of discriminating between males and females (*25*, *72*, *73*). Indeed, sex identity information is widely distributed in the limbic system, including regions upstream of the medial hypothalamus (*17*, *58*, *74*, *75*). Thus, the male- and female-induced neural activation patterns should remain different even without the RBCC. On the basis of these data, we reason that the RBCC deficit–caused changes in social preference likely reflect a decreased interest in certain social targets instead of an inability to discriminate between the sexes.

The RBCC likely promotes and maintains social interest by engaging the mesolimbic dopaminergic pathway (Fig. 3). Dopamine release in the nucleus accumbens (NAc) during initial social encounters has been shown to positively correlate with the total interaction time (*76*). Inhibiting ventral tegmental area (VTA) dopamine neurons decreases social interaction with an unfamiliar social target and the number of nose pokes an animal is willing to emit in order to access a conspecific (*77*–*79*). Conversely, optogenetic activation of the VTA dopaminergic projection to the NAc notably increases social interaction time (*76*, *78*).

The RBCC, especially the MPOA, likely engages the mesolimbic dopaminergic pathway by direct projection to the VTA. Indeed, optogenetic activation of the MPOA-to-VTA terminals evokes dopamine release in the NAc and promotes interaction with pups or potential mates (*38*, *65*). Additionally, activating medial amygdala GABAergic projection to the MPOA increases NAc dopamine release and promotes nose poking to access social targets (*80*). Channelrhodopsin (ChR2)–

assisted circuit mapping revealed that MPOA^{Esr1} cells preferentially target VTA GABAergic cells, suggesting that a disinhibition mechanism underlies the increased dopamine release (*38*). Similar to the MPOA, the VMHvl also promotes social seeking and approach. In male mice, optogenetic activation of VMHvl cells shortens the latency to nose poke for a weaker male intruder, whereas inhibiting the cells has the opposite effect (*49*). Activating female VMHvll^{Cckar} cells promotes social interest in males (*25*). VMHvl activation may evoke dopamine release directly through its glutamatergic projection to VTA dopaminergic cells (VTA^{DA}) or indirectly through its projection to the MPOA (*81*); additional studies are required to identify the exact circuit mechanism.

It is worth noting that the RBCC-VTA^{DA}-NAc circuit is also essential for the reinforcing property of consummatory social actions (Fig. 3). During social behaviors, the level of dopamine increases in the NAc, and animals subsequently develop a preference for the contexts, actions, or specific conspecific cues associated with these experiences (*76*, *82*–*84*). When VTA dopamine release or NAc dopamine receptors are blocked, the social experience-induced reinforcement diminishes (*83*–*86*). Conversely, optogenetically activating the socially relevant cells in the MPOA and VMHvl rapidly enhances the preference of the stimulation-coupled contexts and reinforces the actions that precede the stimulation (*25*, *40*, *65*). Taken together, these results suggest that the increased RBCC activity that is observed during consummatory social actions leads to NAc dopamine increase and reinforces sensory cues and motor actions that are time-locked to the behavior.

Consummatory social actions

Sexual, aggressive, and parental behaviors are recognized on the basis of their stereotyped and distinctive motor patterns. RBCC output to the midbrain, especially the PAG, has been established as a key pathway for the motor execution of social actions (*87*–*90*) (Fig. 3). The VMHvl mainly targets the dorsomedial and lateral PAG, matching the male aggression-induced c-Fos pattern. By contrast, the MPOA mainly targets the ventrolateral PAG, consistent with male mating-induced c-Fos (*81*, *87*, *88*, *90*, *91*). PMv projection to the PAG is relatively sparse (*89*, *90*). Interestingly, the RBCC largely avoids the dorsolateral PAG, a region that is important for defense (*90*). The PAG then relays the hypothalamic signal to spinal cord motor neurons directly or indirectly through the pons and medulla (*92*).

The role of the PAG in female sexual behavior and aggression, which are two VMHvl-mediated social behaviors, has been well established. PAG lesion in rats reduces both behaviors, whereas electric stimulation of the PAG has the opposite effect (*22*, *93*, *94*). More recently, we found that activating the VMHvl to the PAG pathway induces attack in male mice, whereas inactivating the PAG blocks VMHvl stimulation–induced attack (*95*). Interestingly, after PAG inactivation, animals remain highly engaged with the intruder and show attack-like behavior, for example, lunging, but fail to complete the attack sequence (*95*). Consequently, the opponent walks away unharmed. Thus, PAG lesion impairs the motor execution of attack but not the underlying aggressive motivation.

The role of the PAG in male sexual and parental behaviors, which are two MPOA-mediated social behaviors, is less clear. For male mating, c-Fos increased in the PAG (*91*), but large lesions in the PAG accelerated rather than suppressed mounting behaviors (*96*). For parental behaviors, caudal PAG lesion reduces kyphosis, a supine posture during nursing, but does not affect active parental behaviors, such as pup retrieval (*97*). Rostral PAG–lesioned animals still initiate retrieval of pups but have trouble releasing them when held in the mouth (*98*). Optogenetic manipulation of MPOA^{Gal} to PAG projection specifically affects pup grooming (*57*). Thus, different subregions of the PAG are likely involved in different aspects of parental behaviors, and midbrain regions outside of the PAG may drive active parental actions. Indeed, early lesion studies and recent pathway-specific optogenetic activation experiments have suggested that the MPOA engages the ventral pallidum through its projection to the VTA to mediate pup retrieval (*38*, *99*).

How does the RBCC control social actions in broader terms (Fig. 4)? Electrophysiological recordings have shown that medial hypothalamic cells do not carry information regarding specific movements during social behaviors. For example, individual VMHvl cells increase activity during the entire attack sequence regardless of whether the animal is lunging, biting, or tumbling (*100*). Thus, RBCC cells could not instruct moment-to-moment movement; instead, the cells likely serve as a "permissive gate" that allows motor neurons to drive an action based on immediate sensory cues. We speculate that a brainstem–spinal cord reflex circuit exists for each social action, such as biting, retrieving, or mounting, that takes in sensory inputs and activates the local motor neurons to drive coordinated muscle movements (Fig. 4). Visual inputs may enter the circuit at the brainstem level, for example, through the superior colliculus (*101*), whereas spinal cord sensory neurons presumably receive somatosensory inputs. The brainstem–spinal cord circuits, however, are likely unable to operate on their own, either because the motor neurons are under tonic suppression or the inputs from sensory neurons are not sufficiently robust. We suggest that only when the RBCC is activated to remove the inhibition or boost the excitatory drive can the motor neurons respond to the acute sensory inputs. During social encounters, the increased activity from specific cells in the RBCC determines which brainstem–spinal cord circuits are permitted to operate. However, the final release of the action should be triggered by the acute sensory inputs to the brainstem or spinal cord. We speculate that the sensing cells in the brainstem–spinal cord circuits have a limited ability to integrate information from various sensory modalities and thus are activated promiscuously if ungated. Therefore, when the medial hypothalamic cells are artificially activated, the animals initiate social actions toward improper targets (*12*, *13*, *16*). For example, VMHvl activation can induce attacks toward an inflated glove (*16*). Although the glove has no resemblance to a mouse in its smell, sound, or shape, its soft and bouncy surface provides sufficient somatosensory input to trigger the brainstem–spinal cord biting circuit once its gate is opened by artificial VMHvl activation.

A similar conceptual framework for generating lordosis was proposed by Donald Pfaff and

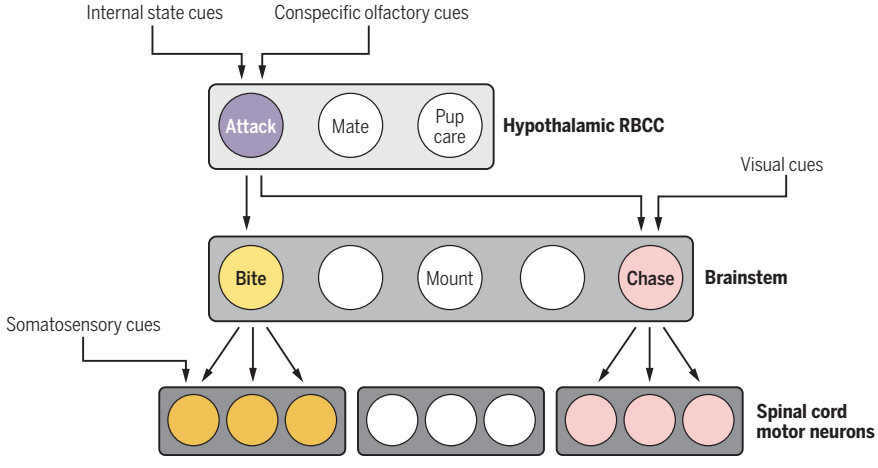


Fig. 4. The RBCC and brainstem–spinal cord dual control of social behaviors. In rodents, the RBCC receives many internal state and conspecific olfactory cues to determine the type of social behaviors to be executed. The RBCC provides permission to specific brainstem–spinal cord cell groups so that those cells can respond to acute sensory visual and somatosensory inputs and generate motor outputs.

K. HOLOSKI/*SCIENCE* BASED ON MEI *ET AL.*

co-workers (*22*). They suggested that a series of brain modules mediates lordosis. The spinal cord modules, coordinated by the brainstem module, control the muscle movements in response to the somatosensory inputs to the spinal cord during male mounting, whereas the output of the hypothalamus module to the brainstem might be essential for ensuring that lordosis only occurs during estrus (*22*). We now know that female VMHvl activity increases during estrus and with the proximity of a male (*23*–*25*, *67*). Thus, the increased VMHvl output provides a window of opportunity for the brainstem–spinal cord circuit to drive lordosis upon receiving the somatosensory cues during male mounting. The dual hypothalamic and brainstem control system ensures that the social action is supported by the animal's physical and reproductive states, is directed toward the right social target (based on hypothalamus input), and happens at the right moment (based on the acute sensory inputs to the brainstem and spinal cord). We speculate that this dual-control system is a common feature in generating innate social actions, and the same principle may apply to learned social actions that vary rapidly with the opponent's behaviors (Fig. 4). As we gain a better understanding of the connectivity between the RBCC and the brainstem–spinal cord circuit that is relevant for each social action, this dual-control system will be further tested and specified in future studies.

Concluding remarks

The RBCC orchestrates all innate social behaviors that subserve reproduction. The MPOA drives male sexual, paternal, and maternal behaviors, whereas the VMHvl and PMv promote female sexual behavior and aggression in both sexes. The MPOA and the VMHvl and PMv may have an antagonistic relationship, as indicated by their opposing roles in multiple social behaviors. At the baseline, the RBCC adjusts an animal's physical activity, hence the chance of social encounters, on the basis of hormonal and metabolic signals. Upon encountering a social target, the RBCC engages the dopamine system to sustain the social interest and reinforce the actions and contexts that lead to the successful completion of social behaviors. The PAG has emerged as a critical midbrain relay for executing VMHvl-driven social behaviors, but the midbrain region that transforms MPOA signals into motor actions remains elusive. Regardless of the exact circuit, we propose a hypothalamic and brainstem–spinal cord dual-control system for the motor execution of each social action. In this model, medial hypothalamus activity determines the broad behavior category based on the animal's internal state and opponent's social identity and opens the gates to allow specific brainstem–spinal cord circuits to respond to the immediate sensory cues and drive moment-to-moment motor output. Though not discussed here, the RBCC also drives a suite of autonomic responses to prepare the body for social actions and triggers neuropeptide and hormone releases that are essential for reproduction after mating (*102*). Lastly, it is worth mentioning that although the RBCC circuit is hardwired developmentally, it remains plastic. The input-output relationship of the circuit can be shaped through experience during development and adulthood, enabling widely different tendencies in the expression social behaviors across individuals (*103*).

REFERENCES AND NOTES

1. L. W. Swanson, *Brain Res.* **886**, 113–164 (2000).
2. S. W. Newman, *Ann. N. Y. Acad. Sci.* **877**, 242–257 (1999).
3. J. L. Goodson, *Horm. Behav.* **48**, 11–22 (2005).
4. L. A. O'Connell, H. A. Hofmann, *J. Comp. Neurol.* **519**, 3599–3639 (2011).
5. J. L. Goodson, D. Kabelik, *Front. Neuroendocrinol.* **30**, 429–441 (2009).
6. J. E. Lischinsky, D. Lin, *Nat. Neurosci.* **23**, 1317–1328 (2020).
7. E. M. Hull, R. I. Wood, K. E. McKenna, in *Knobil and Neill's Physiology of Reproduction*, J. D. Neill, Ed. (Elsevier, 2005), pp. 1729–1824.
8. K. Larsson, L. Heimer, *Nature* **202**, 413–414 (1964).
9. J. R. Moffitt *et al*., *Science* **362**, eaau5324 (2018).
10. J. R. Knoedler *et al*., *Cell* **185**, 654–671.e22 (2022).
11. Y.-C. Wei *et al*., *Nat. Commun.* **9**, 279 (2018).
12. T. Karigo *et al*., *Nature* **589**, 258–263 (2021).
13. D. W. Bayless *et al*., *Cell* **186**, 3862–3881.e28 (2023).
14. H. Lee *et al*., *Nature* **509**, 627–632 (2014).
15. C. F. Yang *et al*., *Cell* **153**, 896–909 (2013).
16. D. Lin *et al*., *Nature* **470**, 221–226 (2011).
17. Z. Guo *et al*., *Neuron* 10.1016/j.neuron.2023.07.011 (2023).
18. J. C. Cavalcante, J. C. Bittencourt, C. F. Elias, *Brain Res.* **1582**, 77–90 (2014).
19. J. C. Cavalcante, J. C. Bittencourt, C. F. Elias, *Physiol. Behav.* **88**, 160–166 (2006).
20. R. A. Ross *et al*., *eLife* **7**, e35960 (2018).
21. A. X. Chen *et al*., *Neuron* **108**, 763–774.e6 (2020).
22. D. W. Pfaff, L. M. Kow, M. D. Loose, L. M. Flanagan-Cato, *Horm. Behav.* **54**, 347–354 (2008).
23. S. Inoue *et al*., *Cell* **179**, 1393–1408.e16 (2019).
24. K. Hashikawa *et al*., *Nat. Neurosci.* **20**, 1580–1590 (2017).
25. L. Yin *et al*., *Neuron* **110**, 3000–3017.e8 (2022).
26. J. Donato Jr. *et al*., *J. Clin. Invest.* **121**, 355–368 (2011).
27. J. Tropp, E. J. Markus, *Physiol. Behav.* **73**, 553–559 (2001).
28. D. Wei *et al*., *Nat. Neurosci.* **26**, 774–787 (2023).
29. Y. Hoshina, T. Takeo, K. Nakano, T. Sato, Y. Sakuma, *Behav. Brain Res.* **61**, 197–204 (1994).
30. B. Powers, E. S. Valenstein, *Science* **175**, 1003–1005 (1972).
31. C. W. Malsbury, D. W. Pfaff, A. M. Malsbury, *Brain Res.* **181**, 267–284 (1980).
32. T. Spiteri, S. Ogawa, S. Musatov, D. W. Pfaff, A. Ågmo, *Behav. Brain Res.* **230**, 11–20 (2012).
33. K. Xiao, Y. Kondo, Y. Sakuma, *Neuroendocrinology* **81**, 56–62 (2005).
34. J. S. Lonstein, G. J. De Vries, *Neurosci. Biobehav. Rev.* **24**, 669–686 (2000).
35. M. Numan, T. R. Insel, *The Neurobiology of Parental Behavior*, Hormones, Brain, and Behavior Series (Springer, 2003).
36. M. Numan, *J. Comp. Physiol. Psychol.* **87**, 746–759 (1974).
37. Z. Wu, A. E. Autry, J. F. Bergan, M. Watabe-Uchida, C. G. Dulac, *Nature* **509**, 325–330 (2014).
38. Y. Y. Fang, T. Yamaguchi, S. C. Song, N. X. Tritsch, D. Lin, *Neuron* **98**, 192–207.e10 (2018).
39. C. Yoshihara *et al*., *Cell Rep.* **35**, 109204 (2021).
40. G. W. Zhang *et al*., *Nat. Neurosci.* **24**, 516–528 (2021).
41. L. Mei, R. Yan, L. Yin, R. Sullivan, D. Lin, *Nature* **618**, 1006–1016 (2023).
42. S. C. Motta *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14438–14443 (2013).
43. A. E. Autry *et al*., *eLife* **10**, e64680 (2021).
44. K. Hashikawa, Y. Hashikawa, J. Lischinsky, D. Lin, *Trends Genet.* **34**, 755–776 (2018).
45. Y. Hashikawa, K. Hashikawa, A. L. Falkner, D. Lin, *Front. Syst. Neurosci.* **11**, 94 (2017).
46. T. Yang *et al*., *Neuron* **95**, 955–970.e4 (2017).
47. R. Remedios *et al*., *Nature* **550**, 388–392 (2017).
48. S. Stagkourakis, G. Spigolon, G. Liu, D. J. Anderson, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25789–25799 (2020).
49. A. L. Falkner, L. Grosenick, T. J. Davidson, K. Deisseroth, D. Lin, *Nat. Neurosci.* **19**, 596–604 (2016).
50. T. Yang *et al*., *Cell* **186**, 1195–1211.e19 (2023).
51. L. Wang *et al*., *Cell Rep.* **26**, 1747–1758.e5 (2019).
52. T. Osakada *et al*., bioRxiv 2022.12.14.519639 [Preprint] (2022).
53. M. Liu, D.-W. Kim, H. Zeng, D. J. Anderson, *Neuron* **110**, 841–856.e6 (2022).
54. S. Stagkourakis *et al*., *Nat. Neurosci.* **21**, 834–842 (2018).
55. S. Stagkourakis *et al*., bioRxiv 2023.02.02.526862 [Preprint] (2023).
56. M. E. Soden *et al*., *Cell Rep.* **16**, 304–313 (2016).
57. J. Kohl *et al*., *Nature* **556**, 326–331 (2018).
58. B. Yang, T. Karigo, D. J. Anderson, *Nature* **608**, 741–749 (2022).
59. A. C. Ribeiro *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16324–16329 (2012).
60. J. E. van Veen *et al*., *Nat. Metab.* **2**, 351–363 (2020).
61. S. M. Correa *et al*., *Cell Rep.* **10**, 62–74 (2015).
62. J. Ryoo, S. Park, D. Kim, *Front. Neurosci.* **15**, 716147 (2021).
63. H. Ye *et al*., *Sci. Adv.* **8**, eabk0185 (2022).
64. S. Musatov *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2501–2506 (2007).
65. J. A. McHenry *et al*., *Nat. Neurosci.* **20**, 449–458 (2017).
66. L. M. Kow, D. W. Pfaff, *Brain Res.* **347**, 1–10 (1985).
67. K. Nomoto, S. Q. Lima, *Curr. Biol.* **25**, 589–594 (2015).
68. Y. He *et al*., *Nat. Commun.* **11**, 2165 (2020).
69. H. Takezawa, H. Hayashi, H. Sano, S. Ebihara, *Front. Med. Biol. Eng.* **6**, 131–137 (1994).
70. W. C. Krause *et al*., *Nature* **599**, 131–135 (2021).
71. K. Fukumitsu *et al*., *Nat. Commun.* **13**, 709 (2022).
72. R. G. Paredes, T. Tzschentke, N. Nakach, *Brain Res.* **813**, 1–8 (1998).
73. R. G. Paredes, M. J. Baum, *J. Neurosci.* **15**, 6619–6630 (1995).
74. J. F. Bergan, Y. Ben-Shaul, C. Dulac, *eLife* **3**, e02743 (2014).
75. D. W. Bayless *et al*., *Cell* **176**, 1190–1205.e20 (2019).
76. B. Dai *et al*., *Cell Rep.* **40**, 111246 (2022).
77. S. Bariselli *et al*., *Nat. Commun.* **9**, 3173 (2018).
78. L. A. Gunaydin *et al*., *Cell* **157**, 1535–1551 (2014).
79. C. Solié, B. Girard, B. Righetti, M. Tapparel, C. Bellone, *Nat. Neurosci.* **25**, 86–97 (2022).
80. R. K. Hu *et al*., *Nat. Neurosci.* **24**, 831–842 (2021).
81. L. Lo *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7503–7512 (2019).
82. S. Krach, F. M. Paulus, M. Bodden, T. Kircher, *Front. Behav. Neurosci.* **4**, 22 (2010).
83. B. J. Aragona *et al*., *Nat. Neurosci.* **9**, 133–139 (2006).
84. B. Gingrich, Y. Liu, C. Cascio, Z. Wang, T. R. Insel, *Behav. Neurosci.* **114**, 173–183 (2000).
85. Y. Xie, L. Huang, A. Corona, A. H. Pagliaro, S. D. Shea, *Neuron* **111**, 557–570.e7 (2023).
86. M. H. Couppis, C. H. Kennedy, *Psychopharmacology* **197**, 449–456 (2008).
87. N. S. Canteras, R. B. Simerly, L. W. Swanson, *J. Comp. Neurol.* **348**, 41–79 (1994).
88. R. B. Simerly, L. W. Swanson, *J. Comp. Neurol.* **270**, 209–242 (1988).
89. N. S. Canteras, R. B. Simerly, L. W. Swanson, *J. Comp. Neurol.* **324**, 195–212 (1992).
90. Z. Jiao *et al*., bioRxiv 2023.05.25.542241 [Preprint] (2023).
91. E. Vaughn, S. Eichhorn, W. Jung, X. Zhuang, C. Dulac, bioRxiv 2022.06.27.497769 [Preprint] (2022).
92. A. A. Cameron, I. A. Khan, K. N. Westlund, W. D. Willis, *J. Comp. Neurol.* **351**, 585–601 (1995).
93. J. Mos, M. R. Kruk, A. M. Der Van Poel, W. Meelis, *Aggress. Behav.* **8**, 261–284 (1982).
94. J. Mos *et al*., *Aggress. Behav.* **9**, 133–155 (1983).
95. A. L. Falkner *et al*., *Neuron* **106**, 637–648.e6 (2020).
96. N. L. Brackett, P. M. Iuvone, D. A. Edwards, *Behav. Brain Res.* **20**, 231–240 (1986).
97. J. S. Lonstein, J. M. Stern, *J. Neurosci.* **17**, 3364–3378 (1997).
98. J. S. Lonstein, J. M. Stern, *Brain Res.* **804**, 21–35 (1998).
99. M. Numan, *Dev. Psychobiol.* **49**, 12–21 (2007).
100. A. L. Falkner, P. Dollar, P. Perona, D. J. Anderson, D. Lin, *J. Neurosci.* **34**, 5971–5984 (2014).
101. D. A. Evans *et al*., *Nature* **558**, 590–594 (2018).
102. L. Yin, D. Lin, *Horm. Behav.* **151**, 105339 (2023).
103. D. Wei, V. Talwar, D. Lin, *Neuron* **109**, 1600–1620 (2021).

ACKNOWLEDGMENTS

Funding: This work was supported by NIH grants R01MH101377, R01MH124927, 1R01HD092596, and U19NS107616 (D.L.) and a basic medical research grant from the Ichiro Kanehara Foundation (T.O.).
Competing interests: The authors declare no competing interests.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 16 May 2023; accepted 25 September 2023
10.1126/science.adh8489

REVIEW

Sleep and the hypothalamus

Antoine R. Adamantidis[1,2]* and Luis de Lecea[3,4]*

Neural substrates of wakefulness, rapid eye movement sleep (REMS), and non-REMS (NREMS) in the mammalian hypothalamus overlap both anatomically and functionally with cellular networks that support physiological and behavioral homeostasis. Here, we review the roles of sleep neurons of the hypothalamus in the homeostatic control of thermoregulation or goal-oriented behaviors during wakefulness. We address how hypothalamic circuits involved in opposing behaviors such as core body temperature and sleep compute conflicting information and provide a coherent vigilance state. Finally, we highlight some of the key unresolved questions and challenges, and the promise of a more granular view of the cellular and molecular diversity underlying the integrative role of the hypothalamus in physiological and behavioral homeostasis.

The hypothalamus is a phylogenetically conserved structure that constitutes only 3% of the volume of the human brain yet is involved in myriad homeostatic functions essential to survival, including thermoregulation, energy and body fluid balance, reproductive or aggressive behaviors, and sleep and wakefulness. Anatomically, the hypothalamus is a federation of cytoarchitectonically defined nuclei containing a large diversity of cell populations that express multiple and distinctive transcription factors, transmitters, neuropeptides, and transmembrane receptors (*1*–*3*). Neurons and glial cells, including ependymal cells and tanycytes, form an intricate network that is embedded in brain-wide structures. This molecular and cellular diversity is consistent with the multiple roles of the hypothalamus in the maintenance of stable physiological set points (e.g., body temperature, energy status, or endocrine secretion) in ever-changing internal and external conditions. A common view is that the hypothalamus contains "sensor-effector" systems that detect deviations of physiological parameters from set points—e.g., through receptors expressed on specialized cells—and trigger responses to keep those parameters within specific ranges. This physiological homeostatic response often precedes the elaboration of goal-directed behaviors ranging from appetite and thirst, reproduction, and parental behaviors to fight-or-flight that represent a behavioral homeostatic response.

Physiological homeostasis drives seemingly antagonistic behaviors. For example, feeding opposes fasting, as fight does to flight responses and sleep does to wakefulness. A consequence of prolonged wakefulness is the accumulation of sleep pressure that dissipates during sleep in a process called sleep homeostasis. Above a set threshold, high sleep pressure triggers "push"/"pull" mechanisms that switch ON sleep-promoting circuits and turn OFF wake-promoting systems in a mutual inhibition network scheme (*4*–*6*). Similarly, neurons controlling food intake and metabolism in the arcuate nucleus and the lateral hypothalamus have long been described as "sensors" and "effectors" because their activity is directly modulated by circulating glucose or the gut hormones leptin and ghrelin, which, in turn, exert brain-wide modulation of intra- and extrahypothalamic circuits that ultimately regulate energy homeostasis through feeding behaviors and tuning of metabolic rate. Similar mechanisms are described for other hypothalamus-based behaviors, including thirst (*7*) or parental behaviors (*8*).

"A consequence of prolonged wakefulness is the accumulation of sleep pressure that dissipates during sleep in... sleep homeostasis."

Despite an apparent anatomical compartmentalization and functional categorization of these homeostatic processes, increasing evidence suggests that single hypothalamic cells, or circuits, overlap in the control of sleep, metabolism, thermoregulation, and energy–fluid homeostasis. Fundamental questions concerning the mechanisms underlying the molecular identity of these cells, their activity pattern, and functional interconnections across homeostatic challenges remain unanswered. This Review illustrates the complexity of the mechanisms at play through the description of cellular networks from the anterior, lateral, and posterior hypothalamus implicated in sleep-wake control and their contribution to ther regulation, food-directed behaviors, and s oscillations.

From oscillations to sleep states

Behaviorally, sleep is defined as a rapidly reversible state of behavioral quiescence and greatly reduced sensory responsiveness to environmental stimuli. Recording of brain and muscle activity using electroencephalography and electromyography, respectively, remains the gold-standard method for defining states of sleep and wakefulness and their transitions (Fig. 1A) (*6*). At NREMS onset, the electroencephalogram (EEG) transitions from a state of high-frequency, low-voltage waves typical of wakefulness to lower-frequency, higher-voltage, slow oscillations (<1 Hz), delta waves (0.5 to 4 Hz), and transient spindles (11 to 16 Hz; Fig. 1A). NREMS is associated with the absence of body movements and decreased body temperature. The transition from NREMS to REMS is signaled by lower voltage, higher-frequency theta oscillations (6 to 9 Hz) and gamma rhythms (>30 Hz) in the EEG, skeletal muscle atonia generated from the pons accompanied by fluctuations of heart beat and breathing rate, and a loss of thermoregulation together with rapid eye movements. Its other name, paradoxical sleep, refers to the contrast between the "wake-like" activity of the brain—although this has been recently revisited (*9*, *10*)—and the concomitant immobility of the body and the limbs. REMS is associated with active dreaming in humans, although dreams are also described during NREMS (*11*).

Current views on the mechanisms of sleep propose that sleep results from mutual inhibitions between sleep- and wake-promoting circuits in subcortical structures (*6*, *12*, *13*), or as an emergent property of oscillatory networks such as the one found in cortical or thalamocortical circuits (*14*), consistent with the local nature of sleep (*15*) (Fig. 1, B and C). These views are not mutually exclusive, and recent findings showed that sleep oscillations such as NREMS spindles are associated with REMS onset (*16*, *17*). Similarly, questions remain as to how local gene expression, protein translation, or their phosphorylation (*18*, *19*) exert pull-push actions on sleep-wake "switches" and ultimately coordinate the cycling of sleep-wake states or sleep homeostasis.

From sleep to "switches"

Empirical mapping of brain circuits that control wakefulness and sleep over recent decades emphasized the hypothalamus as a structure that contains both sleep- and wake-promoting cells, or so-called "switches" in sleep-wake control. Live-cell imaging and perturbation technologies including opto- and chemogenetics in animal models revealed distinctive dynamics of cell populations and their functional

[1]Zentrum für Experimentelle Neurologie, Department of Neurology, Inselspital University Hospital Bern, Bern, Switzerland. [2]Department of Biomedical Research, University of Bern, Bern, Switzerland. [3]Department of Psychiatry and Behavioural Sciences, Stanford, CA, USA. [4]Wu Tsai Neurosciences Institute Stanford University School of Medicine, Stanford, CA, USA.
*Corresponding author. Email: antoine.adamantidis@unibe.ch (A.R.A.); llecea@stanford.edu (L.d.L.)

network at single-cell resolution, deep in the hypothalamus. This led to an improved understanding of the mechanisms governing NREMS, REMS, and wakefulness. This section reviews some of the key experiments that revealed unsuspected dynamics and a large diversity of activity patterns, rarely restricted to a single sleep-wake state.

NREMS

The foundational work of Von Economo *et al.* identified a "sleep center" in the anterior hypothalamus, and the preoptic area (POA) in particular, and soon after, a "wake center" in the posterior hypothalamus (*4*, *6*). This was essential to the modern concept of sleep- and wake-promoting neurons in the brain. Indeed, cell populations selectively active at the onset of

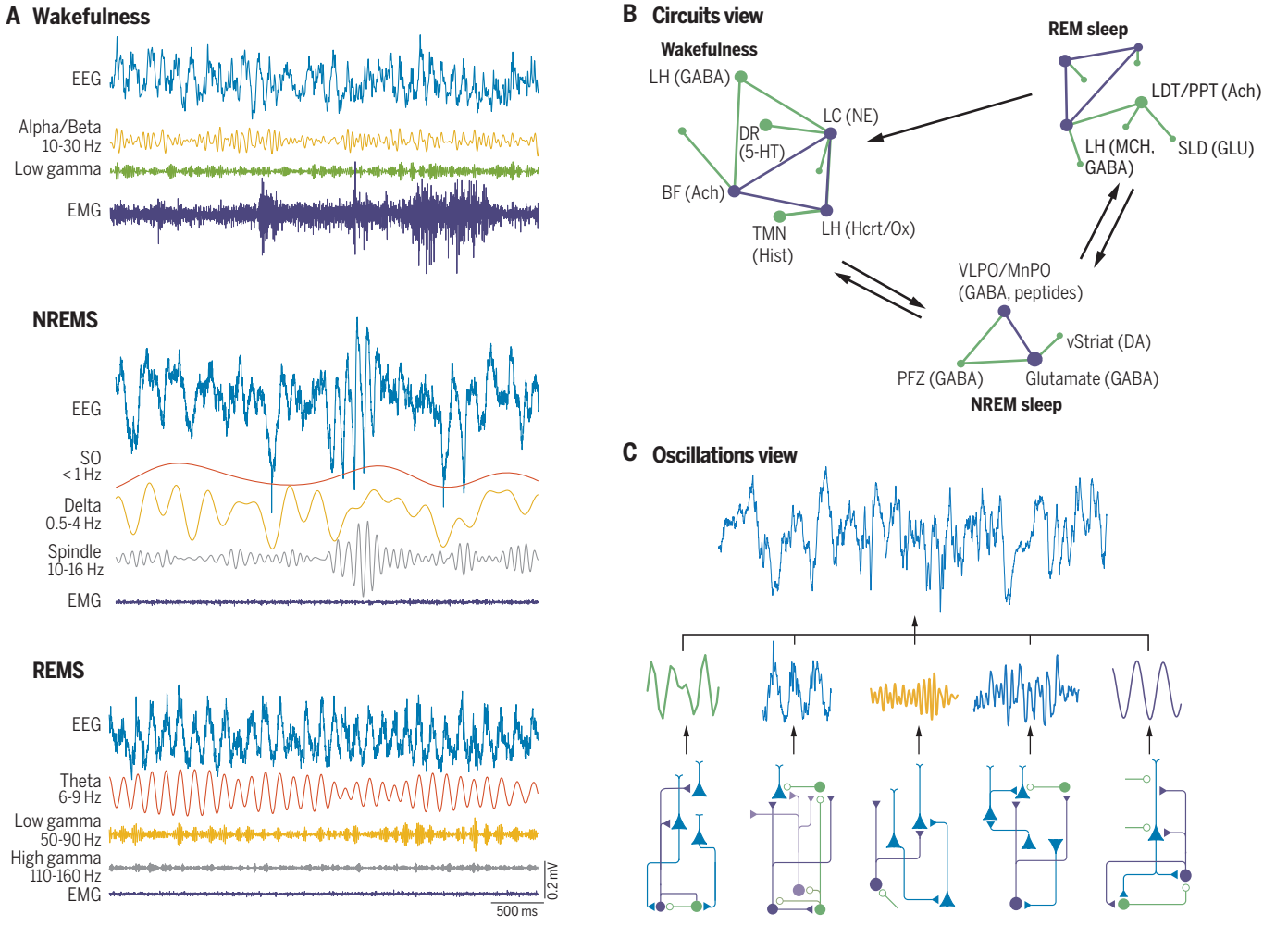


Fig. 1. Cardinal features of a brain EEG during sleep and current models of sleep control. (**A**) Filtering decomposition of EEG signals recorded shows typical state-dependent oscillations in the mouse brain. During wakefulness, the high-frequency, low-amplitude oscillations reflect synchronous activities in cortical and subcortical networks. High EMG activity signals body movement. At the onset of NREMS, EMG activity is low owing to body immobility, and the EEG is dominated by high-amplitude, low-frequency slow wave oscillations (<4.5 Hz) that include slow oscillations (<1 Hz) and delta waves (0.5 to 4.5 Hz) and transient spindles (11 to 16 Hz). The onset of REMS shows predominant EEG theta oscillations (6 to 9 Hz) and gamma rhythms, generated by the septo-hippocampal and cortical networks, respectively. EMG signal is further reduced owing to tonic inhibition of postural muscle. [Adapted from (*98*)] (**B**) Schematic representation of the "circuits view" of the interactions between sleep-promoting areas and arousal systems located in the posterior hypothalamus, the basal forebrain, and the brainstem during the sleep-wake cycle. Wake systems include histaminergic and the hypocretinergic neurons in the hypothalamus, cholinergic neurons in the pons and the basal forebrain, noradrenergic neurons of the LC, and dopaminergic and serotonergic raphe neurons of the brainstem. These cell populations send widespread ascending projections throughout the brain, where they inhibit sleep-promoting systems, and to the entire cerebral cortex. At NREMS onset, sleep-promoting neurons (GABA, galanin) of the VLPO, the median preoptic nucleus (MnPN) located in the preoptic area (hypothalamus), and brainstem parafacial zone (PFZ) progressively inhibit the excitatory influence of wake-promoting systems on thalamic and cortical neuronal populations (*6*). Brain structures responsible for REMS onset and maintenance include cholinergic neurons of the pedunculo-pontine and laterodorsal tegmental nuclei and GABAergic neurons located in the dorsal para-gigantocellular reticular nucleus (DPGi) and vlPAG, cortical and hypothalamic neuronal populations that include neurons producing MCH peptide. [Adapted from (*100*)] (**C**) Schematic representation of the "oscillations view" supporting the observation that sleep can be locally regulated and suggesting that global sleep emerges from the progressive synchronization of local oscillators distributed across the brain. The slow oscillations (<1 Hz) reflect variation in the resting membrane potential of cortical and thalamic neurons that switches between depolarized ("UP") and hyperpolarized ("DOWN") states associated with intense synaptic and bursting activity and cellular quiescence, respectively. Topographically restricted thalamocortical networks from the prefrontal, somatosensory, motor, or visual cortices can act as independent "sleep units" with their own circuit architecture and oscillation pattern.

NREMS and during NREMS (called sleep-active) were eventually identified in the ventral lateral preoptic (VLPO) nuclei, the adjacent median preoptic (MnPO) nuclei, and the basal forebrain using single-cell electrophysiology or imaging, some of which reflect NREMS homeostasis (*6*). Consistent with these correlative findings, chemo- and optogenetic activation of sleep-active neurons in the anterior hypothalamus promoted NREMS and to a lesser extent REMS, possibly through inhibition of wake-promoting neurons of the posterior hypothalamus, whereas their silencing decreases NREMS (*20*). Among the inhibitory neuronal populations in the POA, selective opto- or chemogenetic activation of those coexpressing galanin, cholecystokinin (CCK), corticotropin-releasing hormone (CRH) or tachykinin 1 (TAC1) peptide markers all induce NREMS, but in some case also REMS (e.g., CCK), with variable potency (*2*, *21*) (Fig. 1B). This effect on REMS is due to the existence of cell populations that are more active during REMS than NREMS, as well as REMS-active cells in this area (refer to the section "REMS" below) (*6*). However, wake-active neurons have been also described adjacent to these sleep-promoting neurons that do induce wakefulness upon opto- or chemogenetic activation (*22*). These findings illustrate the large variety of cells and circuits from the anterior hypothalamus that are involved in sleep control.

Consistent with a brain-wide mechanism of sleep, the anterior hypothalamus, and the VLPO in particular, receives direct synaptic inputs from, and projects to, multiple brain regions implicated in sleep-wake control (*6*). Their inhibitory action exerted on wake-promoting neurons in the tubero-mamillary nucleus (TMN), dorsal raphe nucleus and adjacent ventral periaqueductal gray matter (vlPAG), parabrachial nucleus, and locus coeruleus (LC) result in NREMS onset (*23*) (Fig. 1B). Optogenetic activation of sleep-active γ-aminobutyric acid (GABA)–mediated POA neurons suppress firing in wake-promoting hypocretin/orexin (Hcrt/Ox) (*22*, *24*) and histamine-releasing (*2*) neurons.

Thus, accumulating studies implicate an increasing number of NREMS-active and NREMS-promoting cell clusters not only in the hypothalamus but distributed throughout the brain, supporting the view that sleep has multiple origins (*4*, *6*) (Fig. 1B). For instance, NREMS-promoting cells have been described in the lateral hypothalamus (LH) where a subset of inhibitory neurons expressing the long isoform of the leptin receptor showed sleep-promoting properties (*25*) possibly through inhibition of $LH^{Hcrt/Ox}$ neurons (*26*). How these scattered circuits coordinate their response to diverse inputs promoting sleep, and synchronize their outputs to cortical oscillations, remains an unanswered question.

REMS

Unlike NREMS, REMS-active neurons were first identified in the pons where higher activity of cholinergic neurons from the laterodorsal tegmentum (LDT) and the pedunculopontine tegmentum (PPT) coincides with the onset of REMS (*4*, *13*). Additional REMS-promoting cells are located in the ventral medulla and lateral hypothalamus (*13*) (Fig. 1B). Concomitantly, firing of arousal-promoting neurons in the LC, dorsal raphe nucleus, LH, and TMN slows down during NREMS and almost halts during REMS. The fact that manipulating the activity of cholinergic neurons has minimal effect on REMS suggested that other REMS-active cells may be implicated. Rodent models of REMS hypersomnia identified REMS-active cells in the anterior, lateral, and posterior hypothalamus, but also sublaterodorsal nucleus (SLD), the precoeruleus region (PC), the medial parabrachial nucleus, and ventral medulla (*13*). Indeed, diverse populations of hypothalamic cells were found to be REMS-active. A large fraction of LH cells produce the neuropeptide melanin-concentrating hormone (MCH) (*27*) and promote REMS when optogenetically activated, and to a lesser extent NREMS, through multiple downstream pathways, including direct inhibition of wake-active histamine neurons in the TMN and in the brainstem (*28*, *29*). Cells expressing the LIM homeodomain factor Lhx6-positive inhibitory neurons and located in the zona incerta (ZI) dorsal to the LH are activated by sleep pressure (*30*) and promote NREMS and REMS in part by direct inhibition of wake-active Hcrt/Ox and inhibitory cells in the LH (*31*, *32*). Their projection to the supramamillary nucleus (SuM) may contribute to SuM-DG (dentate gyrus) modulation of hippocampus theta oscillations during REMS. Whereas lesions of the extended VLPO (as opposed to its central part) in the anterior hypothalamus predominantly affect REMS (*6*), optogenetic activation of inhibitory cells in this area affects REMS but also NREMS (*2*). This raised the question as to whether NREMS- and REMS-promoting POA cells encompass separate subsets of neurons or whether different mechanisms, including selective neurotransmitter or peptide release or postsynaptic modulation, facilitates one state over the others.

Some of the REMS-active cells in the LH are involved in theta rhythm rather than the control of the onset or maintenance of REMS per se. This is the case for a subpopulation of MCH neurons that modulate the activity of medial septum cells in vitro—the main generator of REMS theta rhythm (*33*)—although the synaptic mechanisms and ultimate effect on REMS architecture or theta rhythm remain to be determined (*28*, *34*). Similarly, a separate population of cells that are MCH-negative but express the vesicular GABA/glycine transporter in the LH (LH^{VGAT}) are strongly activated during REMS (*13*, *35*, *36*), but their causal implication in the onset or maintenance of REMS remains unclear (refer to the section "REMS" below) (*37*). A third population of REMS-active neurons is located in the lateral region of the SuM of the caudal hypothalamus and sends prominent innervation to the dorsal dentate gyrus of the hippocampus. In contrast to their strong activity during REMS, their optogenetic activation modulates theta oscillations and gamma rhythms rather than REMS state duration (*38*), through dual GABA and glutamatergic transmission (*39*). Thus, the original REMS model based on a brainstem mutual inhibition was revisited and now implicates the (lateral) hypothalamus (*13*).

Wakefulness

Early in vivo recordings reported wake-active neurons in the anterior, lateral, and posterior hypothalamus, in particular those producing the Hcrt/Ox peptides that are also active upon arousal stimuli and locomotion (*6*, *36*). Following von Economo's discovery of a sleep center, the LH was identified as a potent arousal center because electrical stimulation of the LH elicits arousal and a cortical EEG resembling wakefulness that is similar to the one induced by electrical stimulation of the brainstem ascending reticular activating system, whereas lesions or inhibition of the LH induce somnolence in rodents, cats, or humans (*36*). Optogenetic activation of $LH^{Hcrt/Ox}$ neurons promotes wakefulness from NREMS and REMS (*40*) through modulation of LC noradrenergic cells (*41*) in a sleep-pressure–dependent fashion (*42*). Likewise, Hcrt/Ox neurons project to wake-active dopaminergic neurons in the ventral tegmental area (VTA) and nucleus accumbens (*43*, *44*). By contrast, lesions or silencing of $LH^{Hcrt/Ox}$, as well as TMN histaminergic, cells do not induce hypersomnolence produced by larger posterior hypothalamic lesions (*45*), but result in unstable boundaries between sleep and wakefulness. These are reflected by fragmentation of both wakefulness and sleep, as seen in narcolepsy or aging (*46*). Unstable boundaries in global sleep states may also be caused by local state heterogeneity in cortical circuits (*47*). Of note, mutations in DEC2, a transcription factor that regulates hypocretin (*Hcrt*) gene expression, result in a need for less sleep, thus raising the question as to whether HCRT activity extends its wake-promoting effect onto longer, possibly homeostatic, timescales (*48*). Similar to the mutual inhibition in brainstem circuits, local inhibition of sleep-promoting cells by wake-promoting cells within the hypothalamus has been proposed; however, such functional connections are sparse (*49*) and remain to be investigated.

LH^{VGAT} inhibitory neurons showed high discharge rates during either wakefulness (45%), in particular goal-directed behaviors such as feeding (*50*), or REMS (40%; <15% are active during NREMS) (*36*), consistent with the high heterogeneity among hypothalamic GABA cells (*1*). Subsets of LH^{VGAT} neurons promote wakefulness from NREMS through direct inhibition of sleep-promoting neurons in the anterior

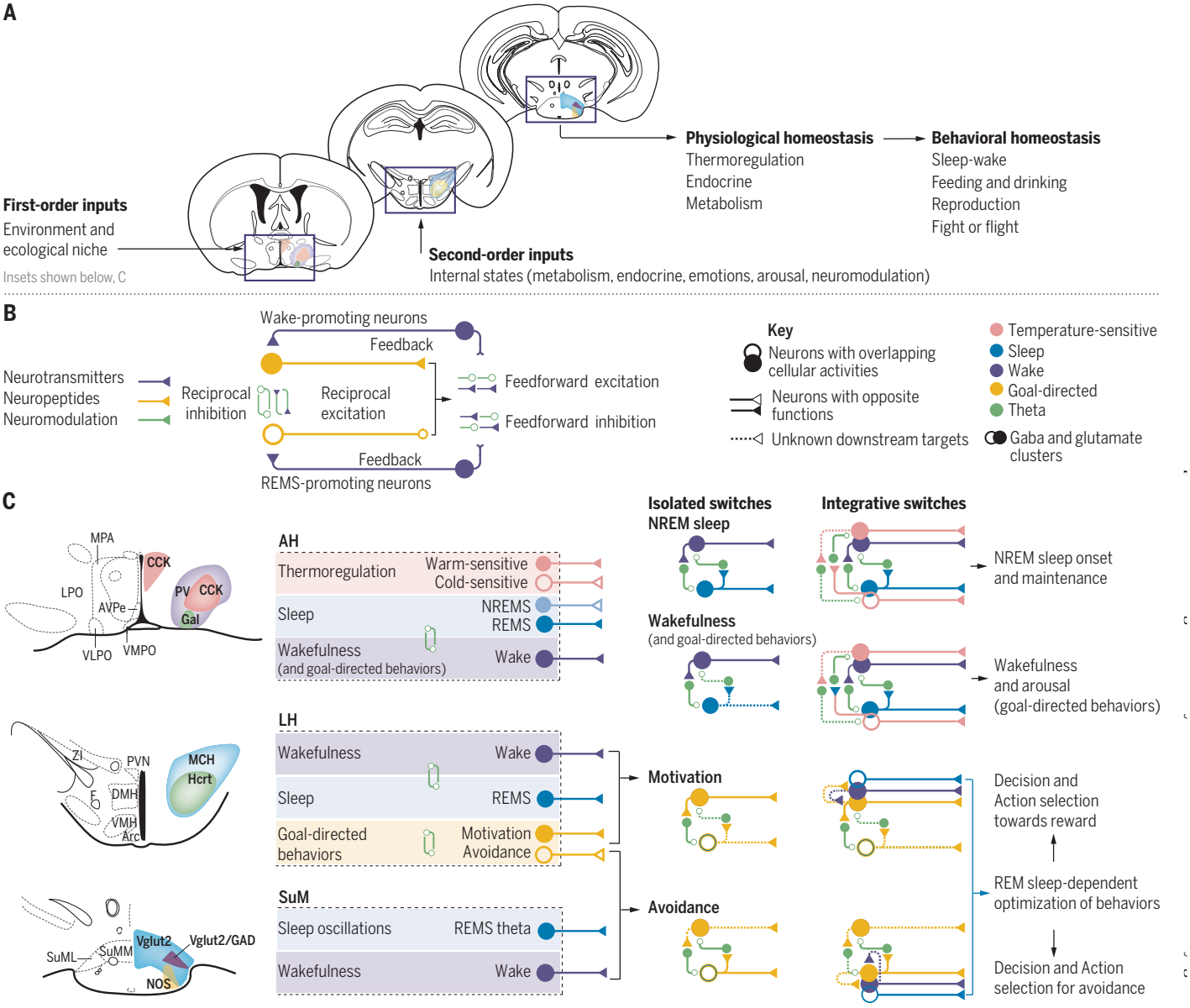


Fig. 2. The spatial organization of hypothalamus homeostasis. (**A**) Schematic representation of the integration of first- and second-order physiological inputs by the anterior, lateral, and posterior hypothalamic circuits and their integrative functions in physiological and behavioral homeostasis. First-order inputs include environmental and ecological factors. Second-order inputs encompass information on the internal states (i.e., homeostatic needs). Neuronal an d non-neuronal networks integrate these inputs into a homeostatic response that adjusts physiological parameters such as body temperature, metabolism, hormonal balance, and appropriate sleep-wake states that often engage a behavioral adaptation and implicate goal-directed behaviors. (**B**) Hypothalamic cells are often considered as state-promoting cells, for example, wake- and REMS-promoting cells represented here based on their activity and ability to drive corresponding states. As "sensor or effector" cells, wake and REMS cells integrate multiple inputs related to sleep-wake needs into coherent outputs that modulate the activity of both hypothalamic and extrahypothalamic circuits and drive a homeostatic response. Similar to neuromodulatory systems, the downstream targets of wake and REMS neuronal outputs are distributed throughout the brain where feedforward inhibition or excitation further drives wakefulness or REMS. Functional feedbacks also participate in the fine-tuning of hypothalamic cell activity during the resetting of homeostasis. (**C**) (Left) Schematic representation of anatomically defined nuclei (left) and some of the cell populations (right) in the anterior, lateral, and posterior hypothalamus. Arc, arcuate nucleus; AVPe: anteroventral periventricular nucleus; CCK, cholecystokinine; DMH, dorsomedial hypothalamus; F, fornix; GAD, glutamic acid decarboxylase; Gal, galanin; Hcrt, hypocretins (also called orexins); LPO, lateral preoptic area; MCH, melanin-concentrating hormone; MPA, medial preoptic area; NOS, nitric oxide synthase; PV, parvalbumin; PVN, paraventricular nuclei; SuMl, supramammillary nucleus, lateral part; SuMM, supramammillary nucleus, medial part; Vglut2, vesicular glutamate transporter 2; VLPO, ventrolateral preoptic area; VMH, ventromedial hypothalamus; VMPO, ventromedial preoptic area; ZI, zona incerta. (Right) Schematics of a hypothalamic circuits supporting physiological and behavioral homeostasis. The POA also contains cells essential for controlling thirst and parental behaviors that are likely to show overlapping pattern of activities. Similarly, the lateral hypothalamus contains cells that regulate wakefulness or REM sleep, and motivation or avoidance behaviors that include feeding or fight-or-flight responses. Wake-promoting cells show overlapping pattern activity with goal-directed cells. These hypothalamic based responses to physiological needs are transformed into decision and action selection through extrahypothalamic circuits in the brain.

hypothalamus (*51*), and disinhibition of thalamocortical networks (*25*), LC, and paraventricular nucleus of the thalamus (PVT) (*25*). The findings that LH^{VGAT} thalamic projections control arousal only from NREMS, whereas stimulation of LC projections wakes animals from both NREMS and REMS, suggest a distributed mode of action through specialized, rather than redundant, arousal pathways. Whether those involved different cell clusters within the LH or a target-specific mechanism remains unclear.

Histamine-releasing neurons from the tuberomammillary nucleus (TMN) represent another wake-promoting circuit in the posterior hypothalamus. Even though genetic deletion of histidine decarboxylase (HDC) (*52*) or its optogenetic manipulation (*53*) induces minimal changes in wakefulness or sleep, in vitro circuit dissection showed that GABA-expressing TMN histaminergic neurons indirectly inhibit sleep-promoting VLPO neurons (*54*) and neocortical cells through putative extrasynaptic modulation (*55*). Lastly, distinct subsets of neurons in the SuM that release GABA or glutamate, or both, are implicated in wakefulness and REMS theta rhythm (refer to the section "The case of REMS theta oscillations" below) (*56*).

It is now widely accepted that NREMS, and REMS, originate from multiple systems distributed across the brain and the body. Yet, a comprehensive model of brain control of sleep is still lacking owing to our partial understanding of the molecular and cellular nature of the sleep homeostat.

Beyond sleep switches

Adding to the complex distribution and functional connectivity of hypothalamic cells, increasing experimental evidence suggests that sleep-wake cells are equally implicated in physiological and behavioral homeostasis, including thermoregulation and awake food-directed behavior, respectively (Fig. 2A). Although this is perhaps not surprising, it strongly contrasts with the prevailing view that single genetically defined hypothalamic cells drive single hypothalamic functions or single behaviors. This section outlines the cells controlling NREMS and body temperature in the anterior hypothalamus, and those implicated in REMS and awake food-seeking behaviors in the lateral hypothalamus. A recent finding implicating the posterior hypothalamus in modulating cell activities in the dentate gyrus during REMS is described as an additional, yet under-investigated, role of the hypothalamus in awake and sleep-dependent modulation of the hippocampus.

The case of thermoregulation and sleep

A locus for "warm-sensitive" neurons, i.e., neurons that increase their firing rate with increasing temperature, was first described in the POA of the anterior hypothalamus (Fig. 2B) (*57*, *58*). By contrast, cold temperature increases the firing of a separate, but intermingled, population of "cold-sensitive" cells (*59*). The POA receives temperature-related information from the primary sensory ganglia in the skin through the dorsal horn of the spinal cord and the lateral parabrachial nucleus. It sends projections to thermoregulatory effector circuits including the dorsomedial hypothalamus (DMH) that controls brown adipose tissue (BAT) thermogenesis through the rostral raphe pallidus (rRPA). Activation of the autonomic POA-DMH-rRPA pathway suppresses thermogenesis and facilitates heat loss through sweating and vasodilation. Local warming or electrical stimulation of the POA causes hypothermia, whereas lesions or pharmacological silencing block the homeostatic regulation of an animal's core body temperature (*58*). Optogenetic or chemogenetic activation of POA^{Vglut2} neurons, or a subpopulation of POA^{LepRb} neurons, is sufficient to induce hypothermia through reduced thermogenesis and locomotor activity. The role of POA inhibitory neurons in thermoregulation remains unclear because optogenetic activation of POA^{VGAT} neurons had no effect on body temperature, but activation of a subset of $VLPO^{VGAT}$ cells induced hypothermia along with a decrease in physical activity. In turn, optogenetic inhibition of these neurons resulted in hyperthermia (*58*, *60*). Whether a mutual inhibition also exists between warm- and cold-sensitive neurons remains unclear (*59*) (Fig. 2B). In addition to the canonical POA → DMH → rRPA pathway, neurons located in other hypothalamic nuclei directly modulate BAT thermogenesis [e.g., (*61*)], but the precise mechanisms and heterogeneity of cell types involved remain to be investigated.

In addition to internal thermoregulation mechanisms and circadian fluctuations of body temperature, homeotherms develop behavioral strategies to ensure a constant temperature by actively seeking sources of warmth in a cold environment, or cold places when exposed to heat. Such homeostatic behavior does not primarily rely on a functional POA but recruits extra-POA circuits and motivation or reward pathways involved in other goal-directed behaviors, including feeding or thirst, to satisfy thermal homeostatic needs, as well as sleep. Indeed, warm temperatures favor sleep, and core body temperature drops before the onset of NREMS, during which it remains low through peripheral vasodilation. By contrast, REMS is associated with an almost complete loss of thermoregulation in many species that is thought to be related to alternative energy allocation during sleep (*62*). Thus, sleep and thermoregulation studies converge toward neurons in the anterior hypothalamus as common substrates for sleep and thermoregulation. Whether sleep and thermoregulation pathways engage overlapping, or separate, neuronal populations remains unclear. Notably, a subset of warm- and cold-activated neurons in the POA show increased firing rate during NREMS and wakefulness, respectively, and POA lesions decreased NREMS and REMS and block homeostatic response to warmth (*63*). Similar to electrical stimulation, warming of the POA induces sleep and enhanced the amplitude of EEG slow oscillations and delta-wave activity, both markers of sleep depth (*63*). Opto- or chemogenetic activation of $VLPO^{Gal}$ promotes sleep and heat loss in mice (*21*, *64*). As an exception, VLPO lesions lead to a decrease in NREMS duration and altered cortical oscillations without affecting body temperature, whereas VMPO lesions have the opposite effect, including no sleep changes and an increase in the circadian amplitude of body-temperature changes (*6*). Similarly, heat-tagged neurons in the medial preoptic area (MPA) revealed two populations driving sleep and thermoregulation, or sleep alone, respectively (*65*). Collectively, these findings emphasize the need for comprehensive approaches to map hypothalamic cells and circuits supporting both sleep and thermoregulation at single-cell resolution.

Notable findings on the mechanisms of anesthesia expand the link between sleep, thermoregulation, and metabolism within POA networks. Experimental reactivation of hypothalamic neurons that were previously activated (and tagged) by dexmedetomidine-induced sedation, or sleep rebound, both increase NREMS in mice (*20*, *64*). These findings support the view that anesthesia and sleep share common pathways in the anterior hypothalamus, although contrasting findings suggest otherwise (*66*). Whether those reflect changes in sleep, sleep-like activity associated with anesthesia, thermoregulation, or bottom-up modulation of the thalamocortical network and ultimately consciousness remains unclear. Furthermore, they raise the question as to what percentage of cells exhibit overlapping activities between these states, and how NREMS and body temperature can be functionally disconnected in experimental procedures.

An additional cell population adjacent to the VLPO and expressing the pyroglutamylated RFamide peptide (QRFP) induced long-lasting hypothermia, hypometabolism, and very-low-amplitude EEG and decreased breathing or heart rate when activated with optogenetics or pharmacogenetics (*59*, *67*). This hibernation-like state relies on GABA or glutamate signaling from the POA to the DMH. These findings further confirmed that torpor-like states share, at least, some of the pathways regulating thermoregulation. Likewise, torpor-inducing TRPM2-expressing neurons in the DMH receive projections from the arcuate and suprachiasmatic nucleus and send projections to brain regions involved in thermoregulation (*68*). How this is integrated with sleep control, brain

oscillations, and metabolic states is an important research endeavor with translational applications in emergency medicine and space travel (*69*).

The case of goal-directed behaviors and sleep

In addition to being implicated in the control of sleep and wakefulness, the LH was once called the "pacemaker for food motivation," which reflects its ability to control both the timing and the generation of goal-directed behaviors toward natural or drug rewards according to homeostatic needs (*70*) (Fig. 2, A and B). Lesions of the LH induce the "lateral hypothalamic syndrome," which is characterized by hypoactivity, somnolence, and ultimately, hypophagia, adipsia, and a global decrease in motivation, whereas electrical stimulation of the LH has the opposite effect and results in "voracious" eating and increased body weight, self-stimulation, and motivation (*70*). This is consistent with a role for the LH and its inhibitory neurons in arousal and awake behaviors, because the activity of LH^{VGAT}, LH^{Vglut2}, and $LH^{Hcrt/Ox}$ is modulated by hedonic and incentive cues associated with both food and taste aversion and also sensory stimuli (e.g., air puff), as well as predation and escape behaviors (*36*, *50*, *70*) (Fig. 2B).

How do LH neurons modulate such a broad range of behaviors? Optogenetic stimulation of non-Hcrt/Ox, non-MCH, or LH^{VGAT} neurons, or their projections in the paraventricular nucleus (PVN) of the hypothalamus, induced a rapid and frequency-dependent increase in both food consumption and reward-seeking behaviors—i.e., low- and high-frequency stimulations driving either consumption or rewarding effects, respectively (*71*)—whereas their optogenetic silencing or genetic ablation results in decreased food intake (*72*). Some of these effects are partly mediated by direct modulation of midbrain VTA neurons (*73*) and likely encode appetitive behaviors and motivation, whereas consummatory responses are presumably driven by LH^{VGAT} terminals in the ventral striatum. By contrast, glutamate neurons in the LH suppress feeding behaviors in food-deprived mice (*72*, *73*), and silencing of their projections to the lateral habenula (LHb) enhances consumption of a palatable reward (*74*), implicating the LH-LHb pathways in negative feedbacks on reward pathways. If LH^{VGAT} and LH^{Vglut2} neurons predominantly control feeding and avoidance behaviors, respectively, $LH^{Hcrt/Ox}$ or LH^{MCH} neurons, which represent a smaller fraction of these excitatory and inhibitory neurons, respectively, only moderately affect food intake and motivation on short timescales (*75*). This is likely due to their antagonistic roles in the modulation of metabolic rate, locomotion, and ultimately body weight over long timescales (*76*). Such temporal dynamics are consistent with long-lasting modulatory actions of peptides on hypothalamus and brain targets as compared to the fast action of neurotransmitters (*77*).

Collectively, these studies questioned the mechanisms and computations underlying the coordination of internal and behavioral homeostatic regulation by the LH (Fig. 2, B and C). If being awake is a sine qua non condition to feed, fight, or reproduce—the wake-active profile of $LH^{Hcrt/Ox}$, LH^{VGAT}, or LH^{Vglut2} is consistent with their involvement in goal-directed behaviors—the function of their activity during sleep is not necessarily obvious. The finding that the activity pattern of feeding-related LH^{VGAT} cells during feeding is reactivated during REMS is thought to reinforce circuit connectivity and ultimately optimize feeding-related behaviors (*37*). Our current understanding of the mechanisms underlying physiological and behavioral homeostasis remains limited to the investigation of a single system within a restricted time window. Taking into consideration the dual role of $LH^{Hcrt/Ox}$ cells in metabolism, reward, and arousal, for instance (*78*), revealed key integrative aspects of hypothalamic control of physiological needs. Equally important is the integration of the "amplitude" of homeostatic challenges. Sleep deprivation of an hour, or overnight fasting, is likely to impose fewer constraints on body physiology than a 6-hour sleep deprivation, or a migration. The cellular and molecular mechanisms underlying such different timescales may range from changes in phosphorylation profiles of ion channels (*18*) to the response to DNA damage associated with sleep. Thus, it will be important to design comprehensive experimental strategies that characterize the activity of single LH cells across various homeostatic challenges at different timescales

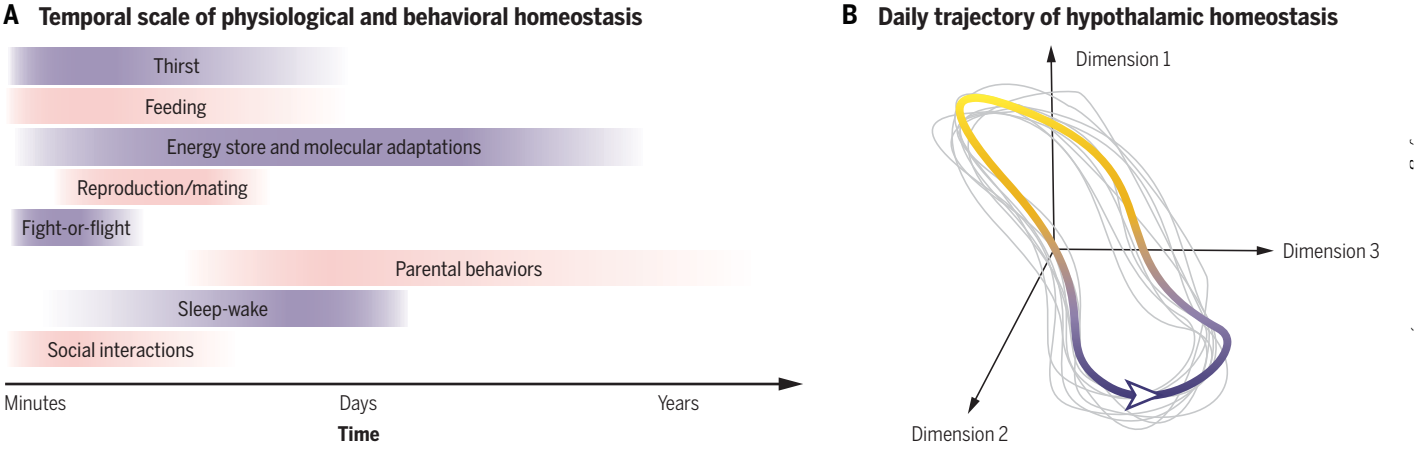


Fig. 3. The temporal organization of the hypothalamus homeostasis. (**A**) Timescale of hypothalamus-based control of physiological and behavioral responses to homeostatic challenges. The activity of hypothalamic cells and circuits across physiological and behavioral homeostasis fluctuates over timescales ranging from fast activity of cells controlling arousal and awakening from sleep to homeostatic or fight-or-flight responses (over seconds to a minute) to slower modulation of physiological homeostasis such as thirst, feeding, or mating that occur on relatively short timescales (from minutes to several tens of minutes). Longer timescales are necessary for molecular components underlying recovery from energy usage or DNA damage in wake-active cells, whereas energy allocation, sleep, and parental behaviors necessitate hours to days, respectively. Although homeostatic processes stem from hypothalamic-based cells, they engage brain-wide networks through a precise coordination of cellular networks across time and space. (**B**) Schematic representation of the trajectories of hypothalamic cells and circuits through recurrent homeostatic routines (gray) are likely conserved (color) in individual organisms across time. For example, energy storage and fluid balance of an animal must be replenished before the preparatory phase, and ultimately onset, of sleep. Larger deviations from set points of energy status, stress due to predation, seasonal behaviors (reproduction, migration), or environmental changes (cold wave, global warming) presumably involve adaptations longer than a day and possibly other systems; the mechanisms driving these homeostatic routines remain unknown.

to decipher the hypothalamic coding of physiological and behavioral homeostasis.

The case of REMS theta oscillations

Hypothalamic cells active during sleep are implicated in other aspects than thermoregulation and goal-directed behaviors. These are presumably involved in modulating brain activity during sleep. This is the case for neurons located in the SuM and their implication in hippocampal oscillations during REMS. The SuM lies dorsal to the mammillary body in the caudal hypothalamus and provides extensive innervation of the cerebral cortex, wake-promoting basal forebrain, and the hippocampus, including the dentate gyrus. Most neurons in the SuM show discharge rates that correlate with hippocampal theta modulation (*79*). Two major populations of neurons that produce and release both GABA (i.e., cells express both glutamate decarboxylase 65, GAD65, and VGAT) and glutamate (i.e., cell express vesicular glutamate transporter 2, vGlut2; $SuM^{GABA/glu}$), or glutamate only, have been described in the SuM that synapse onto dorsal or ventral DG neurons (also CA2 cells; SuM^{glu}), respectively (Fig. 2B). In addition, SuM cells receive reciprocal inputs from the LH, including MCH neurons, DMH, ventromedial hypothalamus (VMH), and VTA (*80*). Whereas SuM lesions reduce the amplitude of theta oscillations during REMS (*39*), chemogenetic manipulations of SuM^{glu} neurons, but not $SuM^{GABA/glu}$, promote wakefulness and increase the amplitude of theta oscillations and gamma rhythms (*56*). Their acute silencing increases NREMS and wakefulness fragmentation and decreases theta oscillations and gamma rhythms. By contrast, $SuM^{GABA/glu}$ neurons activate dorsal DG granule cells and theta-gamma oscillations during REMS (*38*, *39*). Their activation has minimal effect on REMS but induces wake from NREMS, as well as locomotor activity during wakefulness (*38*).

These studies further highlight the complexity of cell identity, in particular the difficulty of determining which neurotransmitters are released by particular cells, and the wide variety of possible postsynaptic control mechanisms. Whether this underlies the dual role of SuM neurons in sleep architecture and brain functions associated with REMS hippocampal oscillations is unknown. Of note, both LH^{MCH} and ZI^{lhx6} neurons have dense projections to the SuM (*81*); however, their role in the regulation of REMS architecture or oscillations remains to be investigated. Recent optogenetic manipulation of CA2-projecting SuM cells selectively during REMS resulted in impaired social memory (*82*), suggesting a possible implication in another aspect of behavioral homeostasis.

In this context, the control of sleep oscillations would be akin to circadian rhythms and neuromodulation whereby thalamocortical or hippocampal networks generate oscillations for cortical communication without the hypothalamus, but transitions between thalamocortical or hippocampal sleep oscillations rely on subcortical structures that include hypothalamic circuits. Thus, clarifying the information carried by SuM neurons to either DG or CA2 reward pathways during wakefulness and REMS likely provides an experimental framework to test the role of hypothalamus modulation in behavioral homeostasis (e.g., social interactions or cognition) and its relationship to sleep. Ultimately, the modulation of sleep oscillations by hypothalamic circuits may establish a previously unknown link between the "circuit" and the "oscillations" views of sleep control (Fig. 1).

Conclusions

Physiological imbalances and changes of internal states trigger motivational drives and goal-directed behaviors in a specific temporal sequence (Fig. 3, A and B). The hypothalamus is central in integrating a variety of thermal, metabolic, endocrine, and environmental stimuli into coherent behavioral responses. The overlap between thermoregulation, goal-oriented behaviors, and sleep described in this Review is not restricted to the example of sleep as suggested by the adaptation of thermoregulation in association with social interaction (hyperthermia) (*83*), the increased arousal in anticipation to food availability (*78*), or the development of feeding strategies under predation (*84*). These recent discoveries emphasize the need for a better understanding of the organization of hypothalamic-based behaviors at single-cell resolution across time and space.

Longitudinal recording of cellular activity in various species (*85*–*87*) and homeostatic challenges including temperature, metabolism, and sleep using genetically encoded voltage, calcium, or transmitter or peptide sensors holds strong promise for revealing the cellular diversity of, and overlap between, hypothalamic neurons controlling body temperature, sleep, or the goal-directed behaviors discussed above.

The genetic identity of hypothalamic cell populations and their neurotransmission remains exceptionally difficult to decipher. Genetically encoded markers are suggestive of a deterministic and evolutionarily conserved neuroanatomical organization, whereas changes in the expression of neuropeptides and neuromodulators, or in the expression of membrane receptors (*88*), including TH (*89*) or MCH (*90*), as well as newly generated neurons (*91*), have been reported in the hypothalamus after sleep deprivation (*92*) or maternal and aggressive behaviors (*93*). Together with the frequency-dependent release of a different set of transmitters or modulators (*40*, *94*), these diverse modes of neurotransmission remain to be investigated across metabolic, temperature, and ecological constraints (e.g., torpor or hibernation). For example, the identification of mirror neurons in the hypothalamus highlights the integration of a social component in hypothalamic-driven behaviors; these may also include social behaviors in anticipation of sleep. At the systems level, resolving the contribution of each of these subpopulations, their dynamics, and effector pathways will refine our understanding of how single neurons share functions as diverse as sleep, thermoregulation, or action selection.

Finally, animals often exhibit preparatory behaviors before sleep that include feeding, drinking, nest building, social huddling, seeking warm places, and curling up to reduce body heat loss. These behaviors implicate brain-wide systems, including the circadian suprachiasmatic nucleus (SCN) (*95*), as well as reward (*44*, *96*) and likely thermoregulatory pathways (Fig. 3, A and B). Accordingly, infusion of glucose in the POA promotes NREMS sleep by increasing the activity of sleep-promoting neurons (*23*), whereas the satiety hormone leptin inhibits $LH^{Hcrt/Ox}$ neurons (*26*), thereby disinhibiting sleep-promoting circuits. Internal state variables that are permissive for sleep—e.g., metabolic signals, circadian synchronization, external and core body temperature, immune status, stress, and emotional state—may generate a multidimensional energy landscape that defines neural trajectories that ultimately trigger sleep, similar to what has been proposed for aggression (*97*). This would be consistent with a prioritization of resource allocations and behavioral strategies associated with the hypothalamic control of multiple homeostats (*62*).

The organization of coordinated physiological and behavioral responses to homeostatic challenges implicates cellular networks beyond the hypothalamus. The stability of these responses across time and space is essential to ensure an optimal adaptation of and survival. A better understanding of those responses from a molecular to a brain-wide perspective will have tremendous consequences for the identification of new treatments for disorders of sleep, metabolism, and brain health in humans.

REFERENCES AND NOTES

1. L. E. Mickelsen *et al*., *Nat. Neurosci.* **22**, 642–656 (2019).
2. S. Chung *et al*., *Nature* **545**, 477–481 (2017).
3. J. R. Moffitt *et al*., *Science* **362**, eaau5324 (2018).
4. F. Weber, Y. Dan, *Nature* **538**, 51–59 (2016).
5. P.-H. Luppi *et al*., *J. Physiol. Paris* **100**, 271–283 (2006).
6. T. E. Scammell, E. Arrigoni, J. O. Lipton, *Neuron* **93**, 747–765 (2017).
7. C. A. Zimmerman, D. E. Leib, Z. A. Knight, *Nat. Rev. Neurosci.* **18**, 459–469 (2017).
8. C. Dulac, L. A. O'Connell, Z. Wu, *Science* **345**, 765–770 (2014).
9. R. Maciel *et al*., *Biochem. Pharmacol.* **191**, 114514 (2021).
10. M. Aime *et al*., *Science* **376**, 724–730 (2022).
11. F. Siclari, G. Bernardi, J. Cataldi, G. Tononi, *J. Neurosci.* **38**, 9175–9185 (2018).
12. F. Weber *et al*., *Nature* **526**, 435–438 (2015).
13. P.-H. Luppi, C. Peyron, P. Fort, *Sleep Med. Rev.* **32**, 85–94 (2017).

14. J. M. Krueger *et al*., *Nat. Rev. Neurosci.* **9**, 910–919 (2008).
15. F. Siclari, G. Tononi, *Curr. Opin. Neurobiol.* **44**, 222–227 (2017).
16. M. Bandarabadi *et al*., *Nat. Commun.* **11**, 5247 (2020).
17. B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas, G. Buzsáki, *Neuron* **90**, 839–852 (2016).
18. A. Kempf, S. M. Song, C. B. Talbot, G. Miesenböck, *Nature* **568**, 230–234 (2019).
19. F. Brüning *et al*., *Science* **366**, eaav3617 (2019).
20. Z. Zhang *et al*., *Nat. Neurosci.* **18**, 553–561 (2015).
21. D. Kroeger *et al*., *Nat. Commun.* **9**, 4129 (2018).
22. K. Prokofeva *et al*., *J. Neurosci.* **43**, 4075–4092 (2023).
23. C. Varin *et al*., *J. Neurosci.* **35**, 9900–9911 (2015).
24. Y. C. Saito *et al*., *Front. Neural Circuits* **7**, 192 (2013).
25. C. G. Herrera *et al*., *Nat. Neurosci.* **19**, 290–298 (2016).
26. P. Bonnavion, A. C. Jackson, M. E. Carter, L. de Lecea, *Nat. Commun.* **6**, 6266 (2015).
27. O. K. Hassani, M. G. Lee, B. E. Jones, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2418–2422 (2009).
28. S. Jego *et al*., *Nat. Neurosci.* **16**, 1637–1643 (2013).
29. R. R. Konadhode *et al*., *J. Neurosci.* **33**, 10257–10263 (2013).
30. C. Blanco-Centurion, A. Vidal-Ortiz, P. Shiromani, *Sleep* **45** (suppl. 1), A57 (2022).
31. K. Liu *et al*., *Nature* **548**, 582–587 (2017).
32. A. Vidal-Ortiz, C. Blanco-Centurion, P. J. Shiromani, *Sleep* zsad217 (2023).
33. R. Boyce, S. D. Glasgow, S. Williams, A. Adamantidis, *Science* **352**, 812–816 (2016).
34. M. J. S. Chee, E. Arrigoni, E. Maratos-Flier, *J. Neurosci.* **35**, 3644–3651 (2015).
35. O. K. Hassani, P. Henny, M. G. Lee, B. E. Jones, *Eur. J. Neurosci.* **32**, 448–457 (2010).
36. B. E. Jones, *Neuropsychopharmacology* **45**, 6–20 (2020).
37. L. T. Oesch *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19590–19598 (2020).
38. F. Billwiller *et al*., *Brain Struct. Funct.* **225**, 2643–2668 (2020).
39. L. Renouard *et al*., *Sci. Adv.* **1**, e1400177 (2015).
40. A. R. Adamantidis, F. Zhang, A. M. Aravanis, K. Deisseroth, L. de Lecea, *Nature* **450**, 420–424 (2007).
41. M. E. Carter *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2635–E2644 (2012).
42. M. E. Carter, A. Adamantidis, H. Ohtsu, K. Deisseroth, L. de Lecea, *J. Neurosci.* **29**, 10939–10949 (2009).
43. S. L. Borgland, S. A. Taha, F. Sarti, H. L. Fields, A. Bonci, *Neuron* **49**, 589–601 (2006).
44. A. Eban-Rothschild, G. Rothschild, W. J. Giardino, J. R. Jones, L. de Lecea, *Nat. Neurosci.* **19**, 1356–1366 (2016).
45. T. Tsunematsu *et al*., *Behav. Brain Res.* **255**, 64–74 (2013).
46. S.-B. Li *et al*., *Science* **375**, eabh3021 (2022).
47. Y. Nir, L. de Lecea, *Neuron* **111**, 1998–2011 (2023).
48. A. Hirano *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3434–3439 (2018).
49. D. Burdakov, M. M. Karnani, *Curr. Biol.* **30**, 4063–4070.e2 (2020).
50. J. H. Jennings *et al*., *Cell* **160**, 516–527 (2015).
51. A. Venner, C. Anaclet, R. Y. Broadhurst, C. B. Saper, P. M. Fuller, *Curr. Biol.* **26**, 2137–2143 (2016).
52. R. Parmentier *et al*., *J. Neurosci.* **22**, 7695–7711 (2002).
53. A. Venner *et al*., *J. Neurosci.* **39**, 8929–8939 (2019).
54. R. H. Williams *et al*., *J. Neurosci.* **34**, 6023–6029 (2014).
55. X. Yu *et al*., *Neuron* **87**, 164–178 (2015).
56. N. P. Pedersen *et al*., *Nat. Commun.* **8**, 1405 (2017).
57. V. V. Feketa, Y. A. Nikolaev, D. K. Merriman, S. N. Bagriantsev, E. O. Gracheva, *eLife* **9**, e55370 (2020).
58. C. L. Tan, Z. A. Knight, *Neuron* **98**, 31–48 (2018).
59. B. A. Upton, S. P. D'Souza, R. A. Lang, *Front. Neurosci.* **15**, 665762 (2021).
60. Z.-D. Zhao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2042–2047 (2017).
61. C. Folgueira *et al*., *Nat. Metab.* **1**, 811–829 (2019).
62. B. Latifi, A. Adamantidis, C. Bassetti, M. H. Schmidt, *Front. Neurol.* **9**, 790 (2018).
63. R. Rothhaas, S. Chung, *Front. Neurosci.* **15**, 664781 (2021).
64. Y. Ma *et al*., *Curr. Biol.* **29**, 3315–3322.e3 (2019).
65. E. C. Harding, N. P. Franks, W. Wisden, *Curr. Opin. Physiol.* **15**, 7–13 (2020).
66. M. Luo *et al*., *Adv. Sci. (Weinh.)* **10**, e2203395 (2023).
67. T. M. Takahashi *et al*., *Nature* **583**, 109–114 (2020).
68. H. Yamaguchi *et al*., Brain-wide mapping of neuronal architecture controlling torpor. bioRxiv 2023.03.03.531064 [Preprint] (2023); https://doi.org/10.1101/2023.03.03.531064.
69. M. Cerri, *Annu. Rev. Physiol.* **79**, 167–186 (2017).
70. G. D. Stuber, R. A. Wise, *Nat. Neurosci.* **19**, 198–205 (2016).
71. M. F. Barbano, H.-L. Wang, M. Morales, R. A. Wise, *J. Neurosci.* **36**, 2975–2985 (2016).
72. J. H. Jennings, G. Rizzi, A. M. Stamatakis, R. L. Ung, G. D. Stuber, *Science* **341**, 1517–1521 (2013).
73. E. H. Nieh *et al*., *Neuron* **90**, 1286–1298 (2016).
74. A. M. Stamatakis *et al*., *J. Neurosci.* **36**, 302–311 (2016).
75. T. Sakurai, *Nat. Rev. Neurosci.* **15**, 719–731 (2014).
76. O. Al-Massadi *et al*., *Nat. Rev. Endocrinol.* **17**, 745–755 (2021).
77. C. Schöne *et al*., *J. Neurosci.* **32**, 12437–12443 (2012).
78. A. Yamanaka *et al*., *Neuron* **38**, 701–713 (2003).
79. R. P. Vertes, *Prog. Brain Res.* **219**, 121–144 (2015).
80. F. Plaisier, C. Hume, J. Menzies, *J. Neuroendocrinol.* **32**, e12829 (2020).
81. S.-G. Oh, Y.-G. Hwang, H.-S. Lee, *Brain Res.* **1748**, 147125 (2020).
82. H. Qin *et al*., *Neuron* **110**, 4000–4014.e6 (2022).
83. D. Liu *et al*., A Hypothalamic Circuit Underlying the Dynamic Control of Social Homeostasis. bioRxiv 2023.05.19.540391 [Preprint] (2023); https://doi.org/10.1101/2023.05.19.540391.
84. I. de Araujo Salgado *et al*., *Neuron* **111**, 2899–2917.e6 (2023).
85. L. C. Leung *et al*., *Nature* **571**, 198–204 (2019).
86. A. L. A. Nichols, T. Eichler, R. Latham, M. Zimmer, *Science* **356**, eaam6851 (2017).
87. S. Liu, Q. Liu, M. Tabuchi, M. N. Wu, *Cell* **165**, 1347–1360 (2016).
88. T. C. Thannickal *et al*., *Sci. Transl. Med.* **10**, eaao4953 (2018).
89. N. Scott, M. Prigge, O. Yizhar, T. Kimchi, *Nature* **525**, 519–522 (2015).
90. J. C. G. Duarte, J. G. P. Ferreira, J. C. Bittencourt, *J. Chem. Neuroanat.* **128**, 102208 (2023).
91. K. Bartkowska, K. Turlejski, P. Koguc-Sobolewska, R. Djavadian, *Neuroscience* **515**, 83–92 (2023).
92. A. Uschakov *et al*., *PLOS ONE* **6**, e16672 (2011).
93. S. Stagkourakis *et al*., *Cell* **182**, 960–975.e15 (2020).
94. C. Schöne, J. Apergis-Schoute, T. Sakurai, A. Adamantidis, D. Burdakov, *Cell Rep.* **7**, 697–704 (2014).
95. B. Collins *et al*., *Neuron* **108**, 486–499.e5 (2020).
96. M. I. Sotelo *et al*., *Curr. Biol.* **32**, 806–822.e7 (2022).
97. A. Nair *et al*., *Cell* **186**, 178–193.e15 (2023).
98. A. R. Adamantidis, C. Gutierrez Herrera, T. C. Gent, *Nat. Rev. Neurosci.* **20**, 746–762 (2019).
99. E. F. Pace-Schott, J. A. Hobson, *Nat. Rev. Neurosci.* **3**, 591–605 (2002).
100. A. R. Adamantidis *et al*., *Sleep* **43**, zsz296 (2020).

ACKNOWLEDGMENTS

We thank all the Tidis and de Lecea Laboratory members, M. Schmidt, and C. Gutierrez Herrera for insightful discussions and comments on previous versions of the manuscript. This manuscript reflects multiple interactions and collaborations that could not be captured in the constrained bibliography. **Funding:** This work was supported by the Inselspital University Hospital Bern (A.A.), the European Research Council (CoG-725850 to A.A.), the Swiss National Science Foundation (310030_188761 and CRSII5_213495 to A.A.), the Synapsis Foundation (2020-PI04 to A.A.), the University of Bern (A.A.), grants from the National Institute of Health (AG08220201, DA05505602, MH12814003, and MH11647005 to L.d.L.), the Knight Initiative for Brain Resilience (KCG-113 to L.d.L.), and BD2 (Breakthrough Discoveries for thriving with Bipolar Disorder to L.d.L.). **Competing interests:** The authors declare no competing interests.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

Submitted 14 June 2023; accepted 8 September 2023
10.1126/science.adh8285

EDITORIAL

First things first

H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org

Although trustees and administrators at universities in the United States are acutely aware of the mental health crisis on their campuses and consistently rate it as one of the most pressing challenges, it usually only gets attention when a tragedy strikes a member of the student body. Those moments illuminate the lack of adequate mental health care for undergraduate and graduate students, but are hastily followed up with meetings admonishing administrators to hire more counselors. Despite the constant drumbeat of these episodes, hardly any universities have prioritized mental health in a comprehensive and proactive way. The newly appointed president of Dartmouth College, Sian Beilock, has taken a refreshing approach in naming campus mental health as the first priority in her inaugural address. It's long past time.

I caught up with Beilock for a wide-ranging conversation about this issue. Our meeting occurred at a time when campuses are strained by the unfolding war in Israel, and that made our discussion all the more poignant. Beilock is a psychologist who has studied the physiological effects of anxiety and stress, and that has informed her priorities. "I've seen people with so much potential not be able to reach it because of stress or anxiety," she said. "And I know the power of equipping students and faculty and communities to deal with these issues."

> "...hardly any universities have prioritized mental health...

Presidential inaugural addresses are typically bland efforts aimed at satisfying multiple constituencies using the latest academic buzzwords interspersed with tributes to the history of the institution. They rarely lead with problems, particularly those involving the campus in question. But Beilock led with the fact that public confidence in higher education in the US has plummeted to 36%, and that poor mental health on campus is a major cause. When I asked her if this aspect of student well-being distracted from a message about the academic work of the university, she refused to concede they were mutually exclusive. "I think that to get the best out of people, they have to feel okay and have the tools to get help when they need it," she said. "And I think sometimes we as principal investigators and leaders don't always make that clear." She said the message should be: "'We care about how you are and how you succeed because feeling well will allow you to succeed even better.'"

Beilock pushed through some common impediments to getting this done. She was quick to point out that the effort had to touch the whole campus. Mental health services are often viewed as a resource for undergraduates and not graduate students, but she has been careful to set it up for all students. It's also common to view the hiring of more counselors as the sole solution, but she sees involvement of the whole campus as equally important. She will have to work hard to get faculty and students to realize that this is something that has to be owned by everyone. Former AAAS (American Association for the Advancement of Science, the publisher of *Science*) leader Alan Leshner has written about this problem and agrees. "Faculty fear dealing with any kind of psychological problems in their students, and generally turn away from the issues out of fear that they'll do the wrong thing," Leshner told me. "So it's important to both train and reassure faculty that they can do this, as long as they don't try to be substitutes for professionals and recognize that their job includes recognizing, sympathizing, and referring students to professionals in this domain."

I asked Beilock what principal investigators or lab members should do when worried about a student coworker. "First, I would understand what resources are available on campus and help students to understand what those are and connect with them," she said. But she also implored faculty to be more open with regard to their own humanity. "I also think it's okay for faculty to talk about their own distress and to let students know that we're all imperfect humans. It often lowers the obstacles to seeking help or getting involved when you understand that your principal investigator also has had some of these same issues."

Beilock's message is that the ideas and discoveries of professors and scholars are important, but students, faculty, and staff are people first. The success of higher education depends on enabling students to fulfill their potential. This rests squarely on mental well-being, and when this becomes a top priority, then perhaps those who have lost confidence in universities will again recognize the positive contribution of these institutions to individuals and society.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.adl4902

NEWS

> "I guess I could say it eases the weight on my heart a little bit, but we have a long ways to go."
>
> **New England Aquarium scientist Philip Hamilton,** to WBUR, on evidence that the decline in population of the critically endangered North Atlantic right whale may be leveling off.

IN BRIEF

Edited by **Jeffrey Brainard**

LEADERSHIP

eLife editor fired for Gaza comments

Michael Eisen, editor-in-chief of the prominent open-access journal *eLife*, revealed this week he lost his job for publicly endorsing a satirical article that criticized people dying in Gaza for not condemning the recent attacks on Israel by the Palestinian group Hamas. Several editors have already resigned to protest his dismissal. In a 13 October social media post, Eisen, a geneticist at the University of California, Berkeley, praised the "moral clarity" of the fake news story by The Onion website. A day later, Eisen, who is Jewish, expressed horror about Hamas's attacks on Israel but also about "collective punishment" of Palestinians in Gaza. Critics called Eisen's endorsement of the satire immoral. They called for him to resign and authors to boycott *eLife*. Some defenders petitioned the Howard Hughes Medical Institute, *eLife*'s publisher, not to punish the editor, citing Eisen's academic freedom of speech. But *eLife* fired him, citing "detrimental" leadership and communication. After taking the helm of *eLife* in 2019, Eisen, a longtime critic of traditional subscription journals, introduced controversial policies, such as ceasing to accept or reject papers; the journal performs peer review of submitted manuscripts for a fee.

Tighter biosecurity rules unveiled

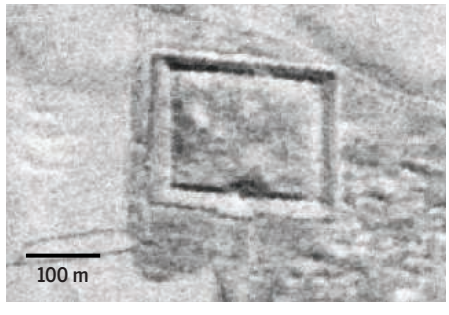
POLICY | To prevent misuse by terrorists, officials have announced stricter guidelines on selling nucleic acids and devices that print genetic material. Voluntary federal guidelines issued in 2010 recommended suppliers block sales to unverified buyers if they could use the products, widely utilized in biomedical research, to construct any of 68 regulated microbes and toxins harmful to humans, animals, or plants. The revisions, released this month by the U.S. Department of Health and Human Services, extend the guidance to cover snippets as short as 50 nucleotides that could increase the toxicity or pathogenicity of organisms not on those lists. The guidelines also ask companies to incorporate automatic screening in desktop devices that print nucleic acid sequences. Some biosecurity experts lauded the changes but say compliance should be mandatory, a policy Congress would have to authorize. The nonprofit Nuclear Threat Initiative is working to create an international organization to codify similar guidelines globally.

NIH nominee gets Senate hearing

LEADERSHIP | Monica Bertagnolli's nomination to lead the National Institutes of Health (NIH) moved forward in the U.S. Senate last week after members grilled her on policy issues, but opposition from a key senator this week has complicated her confirmation. Senator Bernie Sanders (I–VT), who chairs the Senate health committee that held the session, announced he would vote no because he was not convinced she would work to control drug prices. During the hearing, Bertagnolli said she'd work to make NIH-developed treatments "available and affordable" but declined to comment on a proposed "reasonable pricing" clause in contracts with companies. Democrats, with whom Sanders caucuses, hold a one-seat majority on the health panel, so Bertagnolli's nomination will need the support of at least one Republican member to proceed to a floor vote—and none of those senators had endorsed her when *Science* went to press. Bertagnolli, a surgical oncologist, has directed the National Cancer Institute since October 2022. NIH has been without a Senate-confirmed director since Francis Collins stepped down in December 2021.

Spy satellites find Roman forts

ARCHAEOLOGY | Flying over the Syrian desert in the 1920s, a French aviator spotted the square and rectangular outlines of Roman forts. He eventually counted 116, in an arc 1000 kilometers long. For decades, historians assumed they were the remains of a defensive network used to fend off the Sasanian Empire, Rome's rival to the east, starting around 200 C.E. This week in *Antiquity*, archaeologists report a new interpretation. In declassified U.S. satellite imagery from the 1950s and '60s, they found 396 more forts scattered across what is now Iraq and Syria, many far to the west of the border between the ancient empires. Rather than a defensive linear screen like Hadrian's Wall, erected on the Roman frontier in England, the Syrian forts were put in place to protect and tax caravans and traders moving along the Silk Road, the authors argue. Recent development has likely destroyed some forts: Eighty percent of those identified in the old satellite images are no longer



A spy satellite photograph shows a suspected Roman fortification in northern Syria.

IN FOCUS White storks patrol the edge of a controlled fire in Kenya's Maasai Mara National Reserve, hoping to catch prey, such as worms, insects, and spiders, fleeing from the heat. The photo was among those honored in this year's Wildlife Photographer of the Year competition, produced by the Natural History Museum in London.

PHOTO: © ELZA FRIEDLÄNDER, WILDLIFE PHOTOGRAPHER OF THE YEAR

visible from space, says co-author Jesse Casana of Dartmouth College.

EPA targets leaded plane fuel

PUBLIC HEALTH | The U.S. Environmental Protection Agency (EPA) last week declared that the leaded gasoline used in small airplanes threatens public health, paving the way for the government to restrict its use. The leaded fuel contributes about 70% of U.S. air pollution containing lead, a notorious neurotoxin. EPA noted extensive research showing lead causes cognitive impairment, especially in children whose bodies are still developing; the agency says there is no safe level of exposure. EPA said 5 million people, including 363,000 children ages 5 or younger, reside within 500 meters of airports, where studies have shown exposure is higher. Many are members of families with low incomes and are people of color. Lead is added to the fuel of older, piston-driven aircraft to reduce the risk of engine failures. Aviation industry groups and manufacturers are studying lead-free alternatives; health advocates want them to speed up the transition.

A biohub for immune-cell sensors

BIOMEDICINE | Columbia, Yale, and Rockefeller universities have formed a new research partnership that will receive $270 million over the next 10 years to create biosensors by harnessing the natural powers of immune cells to detect disease. The Chan Zuckerberg Initiative (CZI) will provide the bulk of the money, with New York City and New York state each chipping in $10 million. The institute aims to develop methods, still in their infancy, that rely on the ability of immune cells to recognize and remember signals from diseased cells and organs. In theory, blood samples containing sensor cells that scientists have engineered could reveal specific cancers, neurodegenerative diseases, and other illnesses. This becomes the fourth "biohub" project funded by CZI, a philanthropy created by Meta founder Mark Zuckerberg and pediatrician Priscilla Chan, who are married.

Excel fixes science formatting

SOFTWARE | Microsoft has modified its Excel spreadsheets to allow users to turn off automatic-formatting functions that have for years frustrated scientists because they introduce errors or otherwise mar entries about research. One of these features automatically converts text to calendar dates—for example, Excel has rendered SEPT2, a common abbreviation for the gene *Septin-2*, as 2 September. One in five genetics papers in top scientific journals contained errors from the program, a 2016 study found. In a blog post last week, Microsoft said Excel will now display a warning message when it detects that at least one of the automatic data conversions is enabled, and the options for disabling them will be more visible in Excel's menu (at File > Options > Data > Automatic Data Conversion). The options, piloted last year, will only be available to users of Windows version 2309 and Mac version 16.77 and later ones.

China plans big neutrino catcher

ASTRONOMY | China plans to build one of the world's largest neutrino detectors deep underwater to catch the elusive messengers from deep space. The Tropical Deep-sea Neutrino Telescope (TRIDENT), announced this month, would be operational by 2030. Neutrinos are byproducts of nuclear reactions in stars. They rival photons in number, but rarely interact with ordinary matter. To observe them, researchers place light detectors in a large mass of water or ice and look for the rare flashes each time a neutrino hits an atom. TRIDENT would use 24,000 detectors moored at the bottom of the South China Sea, 3500 meters below the surface, to examine a space measuring 7.5 cubic kilometers. That collection volume is about the same as is planned for an upgrade to the U.S. IceCube Neutrino Observatory at the South Pole, which will make both new instruments far more sensitive than existing ones.

SCIENTIFIC FACILITIES

DOE picks lab to manage coming data tsunami

Billions of gigabytes from accelerators and x-ray sources could swamp national labs

By **Adrian Cho**

At risk of drowning in the output from its own facilities, the Department of Energy (DOE) plans to build a major new computing center specifically to crunch experimental data. The agency will construct a $305 million High Performance Data Facility at the Thomas Jefferson National Accelerator Facility (JLab) in Newport News, Virginia, officials announced last week. To be completed by 2028, the new data center marks an evolution in the agency's approach to science and gives a new lease on life to JLab, a small lab that currently focuses on nuclear physics and accelerator development.

"This is a major response to the growing demand," says Asmeret Asefaw Berhe, director of DOE's Office of Science. DOE's x-ray and neutron sources and atom smashers serve thousands of experiments probing everything from the origins of matter to the structures of materials and proteins. As the intensity of the sources has increased, they produce data ever faster and in greater detail, threatening to overwhelm computing resources at the 10 national laboratories run by the Office of Science. "The need for managing data is only growing," Berhe says.

For example, DOE runs four circular electron accelerators, or synchrotrons, that generate x-rays, and since the 1970s the intensity of such sources has increased 100 billion–fold. And next spring, Argonne National Laboratory in Illinois will finish upgrading its Advanced Photon Source (APS) to boost its intensity by a factor of 500.

Brand new data gushers have arrived, too. In 2009, researchers at SLAC National Accelerator Laboratory in California turned on the world's first x-ray laser, the Linac Coherent Light Source (LCLS), a device driven by a linear accelerator that produces x-ray flashes 1 billion times brighter than those from a synchrotron. SLAC just completed an upgrade to enable the LCLS to produce 1 million flashes per second, increasing its data rate by a factor of 8000.

Within a decade, the data torrent from the five x-ray sources alone should grow by more than a factor of 100 to reach 1 billion gigabytes per year, says Jana Thayer, director of the LCLS's data systems division. That's the equivalent of 1.5 million full-length movies per day, she says. "It's not just that we need to watch these movies," she says. "We need to look at every single frame of every movie and analyze it in real time."

For example, researchers at the LCLS can determine a protein's structure by dropping micrometer-size crystals of it into the ultra-intense beam and detecting the x-rays that scatter from the microcrystals even as they are vaporized. Increasingly, biologists want to catch the protein as it bends or breaks in response to a flash of laser light. To do that, researchers have to tune the timing between the laser flash and the LCLS's x-ray blast to within a millionth of a nanosecond, best done by analyzing the data as it's gathered.

Similarly, experimenters at Argonne can trace with near–atomic-scale resolution the internal structure of a microchip, for example to look for faults. With the upgraded APS, they will be able collect those data in a couple of days, says Nicholas Schwarz, lead for scientific software and data management at the APS. But they would need a month to analyze them, Schwarz says. "That means you're flying blind as you take the data, so we're looking to accelerate the reconstruction."

The JLab facility aims to do just that. Like the supercomputers DOE runs at Oak Ridge National Laboratory in Tennessee and at Argonne, it will include hundreds of thousands of processors and a huge, automated repository of tapes. But the existing supercomputers are optimized to grind away steadily on enormous simulations of the climate or exotic materials or to train artificial intelligence programs. Time on the machines is apportioned months in advance. The new

The Thomas Jefferson National Accelerator Facility, a nuclear physics lab, will host a data hub.

facility will be designed to respond on the fly to experiments' fluctuating demands, says Amber Boehnlein, JLab's associate director for computational science and technology. "If researchers need real time feedback, they have to have access to the computing cycles."

DOE has already taken steps in this direction. It has a private network connecting the science labs. And since 1996, Lawrence Berkeley National Laboratory (LBNL) in California has run the National Energy Research Scientific Computing Center, which serves remote experiments and has even demonstrated real-time analysis of data from a single LCLS experiment. But the new data facility will automate and expand the process, using an improved network to shuffle data among the labs to wherever computing resources are free, Boehnlein says.

The center also marks an inflection point for JLab. Work there has centered on the lab's flagship Continuous Electron Beam Accelerator Facility (CEBAF), which since 1994 has probed the structure of atomic nuclei and individual protons and neutrons. DOE plans to operate CEBAF for at least another decade, but exactly how long is unclear.

In 2020, DOE tapped Brookhaven National Laboratory in New York to build a higher energy machine, the $2.4 billion Electron-Ion Collider (EIC), which will also probe protons and neutrons. JLab physicists would still like to upgrade CEBAF, but many are already working on the EIC, says Gail Dodge, a nuclear physicist at Old Dominion University who led a recent long-range planning exercise for DOE's nuclear science advisory committee. "Even at JLab the highest priority is the EIC."

The new facility helps secure JLab's future. It "diversifies our scientific mission in an area that is so critical and exciting right now," says Stuart Henderson, JLab's director. "It couldn't be better." Berhe emphasizes that JLab won the data hub on the strength of its proposal in a competition among the labs. Still, she adds, "It will clearly contribute to [JLab's] vibrancy for many, many years."

The facility also embodies the continuing evolution of DOE's entire science effort, says Michael Witherell, director of LBNL, which will partner with JLab in the effort. Since its inception in 1977, DOE has built scientific facilities based on particle accelerators. More recently, it has expanded into providing supercomputers. Now, it aims to pioneer weaving together supercomputing and experiments, Witherell says. "We have a chance to be first in the world at showing how all these facilities can work together." ■

GLOBAL HEALTH

First malaria vaccine slashes childhood deaths

Final evaluation brings good news from RTS,S rollout

By **Meredith Wadman**

In a major analysis in Africa, the first vaccine approved to fight malaria cut deaths among young children by 13% over nearly 4 years, the World Health Organization (WHO) reported last week. Data from a pilot rollout of the vaccine, called RTS,S or Mosquirix and made by GSK, also showed a 22% reduction in severe malaria in kids young enough to receive a three-shot series. Hundreds of thousands of children are born annually in the parts of Ghana, Kenya, and Malawi included in the analysis, for which WHO revealed the final data at the annual meeting of the American Society of Tropical Medicine and Hygiene.

"The RTS,S malaria vaccine is already saving lives," said John Tanko Bawa, director of malaria vaccine implementation at the nonprofit PATH. He added, "What we have seen is a considerable impact of a vaccine described as having modest efficacy." The 13% drop in deaths is so remarkable that "I was surprised I didn't hear any gasps when it was stated," joked medical epidemiologist Mary Hamel, who led the WHO pilot program.

> **"What we have seen is a considerable impact of a vaccine described as having modest efficacy."**
>
> **John Tanko Bawa,** PATH

The mortality decline could translate to tens of thousands of lives saved if RTS,S, which WHO approved for widespread use in 2021, is more broadly deployed: In 2021, malaria killed an estimated 468,000 children younger than age 5 in sub-Saharan Africa. Seventeen countries in the region have already won approval to receive doses, which will start to roll out next year. "This was a very large, very robust evaluation done in a real-life setting, and you're seeing this huge impact," says Kwaku Poku Asante, a physician and epidemiologist who directs the Kintampo Health Research Centre and oversaw the analysis in Ghana.

In trial results published in 2015, RTS,S showed 36.3% efficacy against clinical malaria 4 years after toddlers were vaccinated. In the $70 million pilot, launched in 2019, nearly 2 million children ages 5 months to 24 months have been vaccinated in the three countries, enabling researchers to document the real-world effects of RTS,S and study the impact of its rollout on routine childhood vaccinations.

To calculate mortality, the researchers employed tens of thousands of community reporters who conducted household surveys of deaths in kids younger than age 5. They then compared the death rates between children age-eligible to receive three doses of the vaccine and those who were not, in both RTS,S areas and unvaccinated areas. The comparison, covering 46 months, revealed the 13% decline in mortality—excluding accidental deaths—attributed to RTS,S. The researchers used the same method to detect the 22% decline in severe malaria, counting admissions with severe disease at designated "sentinel" hospitals.

Matthew Laurens, a malaria vaccine researcher at the University of Maryland School of Medicine, calls the results "fantastic news" and theorizes that beyond preventing malaria, the RTS,S vaccine may be broadly "training" the immune system, providing a protective benefit against other infections. Measles and tuberculosis shots have shown such general survival benefits.

Using RTS,S did not hurt uptake of other childhood vaccines or reduce bed net use. But some public health leaders worry its use will require trade-offs. "We are still struggling to implement the tools that we [already] have in many countries," David Walton, U.S. global coordinator for the President's Malaria Initiative, told the panel that presented the data last week, noting that the extra cost of RTS,S, at $10 a dose, "is formidable for many countries." A second malaria vaccine called R21 won WHO authorization earlier this month and is likely to be available more cheaply and in greater quantities than RTS,S.

The lengthy, expensive RTS,S pilot program "did contribute to a delay in widespread use of the vaccine," Hamel acknowledged. But without it, "I really believe that questions would have lingered" about the vaccine's effectiveness and impact—and the feasibility of reaching kids. ■

BIOLOGY

‘Why can’t you make a coral out of an anemone?’

By adding coral genes to a nonmineralizing relative, team aims to show how reefs are built

By **Christie Wilcox**

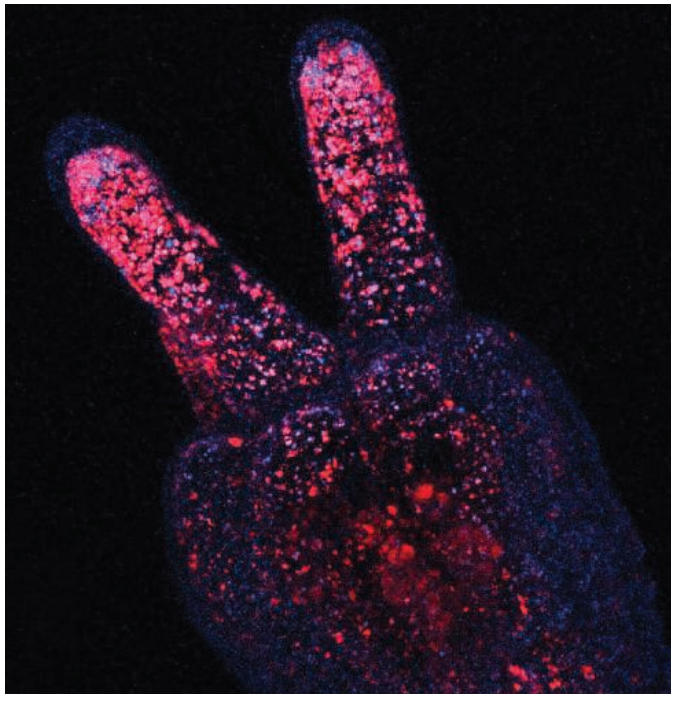
Reefs are vibrant, living structures laid down over time by tiny tentacled animals. But how exactly corals construct the crystals that become the reefs’ craggy rocks—a process known as biomineralization—has long been a mystery. In June, University of Florida (UF) marine biologist Federica Scucchia believed she and her colleagues had finally made a breakthrough. Peering through a microscope at blue and red blobs along the tentacle of a 0.6-millimeter anemone, a soft-bodied coral relative, Scucchia saw signs that the team had managed to coax the creature to produce a coral protein that concentrates reef-building calcium. “I was like, ‘Wow! This cannot be!’” she recalls.

The finding, reported on 5 October in a bioRxiv preprint, has excited coral researchers. Anemones are far easier to manipulate in the lab than corals, which could enable researchers to dissect the workings of biomineralization. That could improve their ability to predict how climate change will affect critical reef ecosystems. “For over 10 years now, I have been trying to find a way to use an invertebrate as a model system for biomineralization studies,” says Tali Mass, a coral physiologist and biomineralization expert at the University of Haifa. The new result marks progress toward that goal, she says.

Scientists have glimpsed some elements of the chemical feat by which corals pull calcium and carbonate ions from seawater and fuse them into reef-building minerals. They know that, in special compartments, corals concentrate ions along with proteins and other molecules to make a slurry known as calcifying fluid. By controlling the contents of this fluid, the corals can encourage the minerals to crystallize into calcium carbonate.

Scientists have also identified many of the genes thought to be involved in this process. But pinning down their actual roles in mineralization has proved to be a nightmare. Corals are finicky and difficult to breed in the lab, explains UF developmental biologist Mark Martindale, making them “a pain in the ass” for genetic engineering.

Anemones, however, are a breeze to work with, and scientists have created genetically engineered anemone lines like those made in mice and other model organisms. Years ago, Martindale was studying the development of a small anemone species called the starlet sea anemone (*Nematostella vectensis*) and had gotten “pretty bored of it.” He had a wild thought: “Why can’t you make a coral out of an anemone?”



A genetically engineered anemone produces a coral protein (red) that is concentrating calcium (blue).

After all, anemones and corals are fairly closely related—they’re both Cnidarians—points out Martindale’s laboratory technician, Brent Foster. Anemones share with corals some of the same genes that have been previously implicated in biomineralization. What anemones notably lack are genes for so-called intrinsically disordered proteins, which play crucial roles in accumulating calcium and carbonate ions.

Could an anemone be engineered to make these proteins? To find out, Martindale and colleagues injected the gene that codes for a disordered protein called SpCARP1 directly into *Nematostella* embryos. A fluorescing red molecule indicated the gene was active. The anemone was indeed making the protein.

But was it actually concentrating calcium? Scucchia drew on experience with corals to suggest adding a different fluorescent molecule, Calcein Blue, that glows in the presence of calcium. Sure enough, the same regions making SpCARP1 also seemed to be aggregating calcium. Although many more changes will likely be required before the anemone will create carbonate crystals, this is an important first step, Scucchia says. Foster adds that the team was “pretty excited” when it saw that. “It took us a long time to get there.”

It’s a “really exciting” result, says University of Rhode Island molecular ecophysiologist Hollie Putnam. She adds that scientists can use this basic model to better understand biomineralization by corals, as well as how the process could be affected by environmental change. Rising water temperatures, acidification, and other environmental factors are already taking a toll on reefs around the world. With engineered anemones, researchers could uncover which versions of genes might make reefs more resilient in the face of likely future conditions.

However, Pupa Gilbert, a biophysicist and geobiologist who studies biomineralization at the University of Wisconsin-Madison, says the preprint is “interesting” but she’s “lukewarm” about its significance to understanding biomineralization. She’d prefer researchers focus on corals directly, for example by knocking out specific genes and observing how their absence affects mineralization.

For Martindale, Scucchia, and colleagues, the next step will be to add coral proteins one by one to anemones to try to get them to create actual calcium carbonate crystals—and possibly the minerals of other animals, such as those in echinoderm spines or mammalian tooth enamel.

“We’re trying to push the envelope and see what we’re able to do,” Foster says. ■

PHOTO: CUI JUN/BEIJING YOUTH DAILY/VCG VIA GETTY IMAGES

Project NextGen aims to bring nasally sprayed CO 19 vaccines, already available in China, to the U.S.

COVID-19

Billions boost next-generation weapons against SARS-CoV-2

U.S. government project bets big on intranasal vaccines and monoclonal antibodies, but omits more ambitious approaches

By **Jon Cohen**

Six months after announcing the $5 billion Project NextGen to develop treatments and vaccines that can "stay ahead of COVID-19," the U.S. government has awarded 20 contracts that reveal what much of that sum will support. The Department of Health and Human Services (HHS) announced on 13 October that up to $1.2 billion will go to three vaccinemakers aiming to develop better shots, adding to $1 billion already awarded to test them in yearlong, 10,000-person clinical trials. Another half-billion will support development of monoclonal antibodies (mAbs) that can block SARS-CoV-2 infection.

Project NextGen is far from the all-hands-on-deck effort of the $18 billion Operation Warp Speed, launched in May 2020, that led to the first COVID-19 vaccines in record time. But Jennifer Nuzzo, director of the Pandemic Center at Brown University School of Public Health, says the program could "improve upon existing COVID vaccines and treatments [and] lead to discoveries that can help in our fight against other infectious diseases."

Others say it will do too little to prepare the world for the next pandemic. "The worst thing we could do is leave people with the sense that this is going to be the solution to the future," says Michael Osterholm, an epidemiologist at the University of Minnesota who was on a steering group of researchers that wrote a detailed R&D roadmap for vaccines that could protect against many coronaviruses, not just SARS-CoV-2. But, he adds, "It's a start."

Current COVID-19 vaccines only prevent symptomatic disease for a few months and are even less effective at preventing infection. Among the new awardees hoping to do better are CastleVax and Codagenix, which have developed vaccine candidates that will be sprayed into the nose. The goal is to stimulate immunity at the mucosal membranes, a strategy that might prove better at thwarting SARS-CoV-2 than current shots. A third vaccinemaker, Gritstone bio, uses an injected formulation designed to protect against a broader range of current and future SARS-CoV-2 variants.

Each company will receive $300 million to $400 million but will only be paid in full if they hit milestones such as meeting regulatory requirements for the clinical trials. Florian Krammer, a virologist at the Icahn School of Medicine at Mount Sinai who co-invented the CastleVax vaccine, applauds the decision to award the money to startup companies pursuing the riskiest, most innovative research, rather than the Big Pharma vaccinemakers that Warp Speed favored. "It's going to allow us to move forward," Krammer says.

In contrast to the vaccines from Moderna and the Pfizer-BioNTech collaboration, which consist of strands of messenger RNA (mRNA), CastleVax and Codagenix use live viruses that can infect cells in mucosal membranes, which then produce SARS-CoV-2 proteins. CastleVax stitches the gene for SARS-CoV-2's surface protein, spike, into Newcastle disease virus, which does not harm humans. Codagenix uses a version of SARS-CoV-2 it has engineered to not cause disease, but still stimulate the immune system.

Gritstone relies on mRNA. But unlike the mRNA in current vaccines, which codes for spike, Gritstone's codes for both spike and regions of the virus that remain conserved across a wide range of coronaviruses. The goal is to trigger both antibodies specific to spike and T cells that work against a broad range of variants. The mRNA in the Gritstone vaccine also makes copies of itself in human cells after being injected, potentially producing large amounts of these viral pieces to generate a stronger immune response.

Whereas vaccines train the immune system to fight infections, mAbs do the work themselves, and a subset of NextGen awardees hope to recreate the success mAbs had early in the COVID-19 pandemic. Back then, the U.S. Food and Drug Administration authorized one mAb to prevent infection against SARS-CoV-2 and eight others to treat mild and moderate COVID-19 disease. But viral variants soon dodged the mAbs, and all are now off the pharmacy shelf.

Regeneron will receive the lion's share of the money, $326 million, to develop a better mAb to prevent infections. The company says its candidate binds to a conserved region of the virus, reducing the risk that variants will escape its powers. ModeX Therapeutics—whose scientific leaders include former National Institutes of Health Director Elias Zerhouni—is developing antibodies that aim to thwart variants by targeting several viral sites. Finally, Vir Biotechnology will attempt to deliver mAbs via mRNA technology, injecting the code for each antibody rather than the molecule

itself. That would allow a simple shot of mRNA to take the place of a complex infusion, and make it easier to redesign the antibodies for a newly circulating variant.

Myron Cohen, an infectious disease specialist at the University of North Carolina who advises several companies that make mAbs, notes the drugs might provide better protection against infection than existing vaccines. "MAbs given every 6 months are just like vaccines except there's no wishful thinking required," Cohen says.

But others note that for now, mAbs are far more expensive than vaccines, limiting their use. "If you were investing in new production methodologies for rapidly producing inexpensive antibodies, that makes some sense to me," says Mark Dybul, a former director of the Global Fund to Fight AIDS, Tuberculosis and Malaria. "But if you're just making more antibodies that are going to cost $8000 an infusion, that's not going to do it."

Dawn O'Connell, HHS's assistant secretary for preparedness and response, says her team expects clinical trials for mAbs to begin this fall and tests of the vaccine candidates to start in the winter. O'Connell says the goal is "strengthening us for whatever the COVID-19 virus brings next."

The remaining half-billion dollars in the NextGen funding announced so far will support development of innovative technologies such as skin patches to deliver vaccines, antiviral drugs that might work against a wide range of coronaviruses, and wearable sensors to detect infections. But Osterholm and others note that one ambitious approach is missing: vaccines that could protect against all future variants of SARS-CoV-2 as well as viruses in the same family that have yet to jump into humans. Other entities, such as the nonprofit Coalition for Epidemic Preparedness Innovations and the National Institute of Allergy and Infectious Diseases, have committed $250 million to several companies and academic groups to spur development of these pancoronavirus vaccines.

Few of those products have moved far enough in the development pipeline to earn a place in NextGen. And some experts see this as evidence that support is waning for higher risk R&D that could pay off the next time a pandemic virus emerges. "We're in worse shape now than we were at the beginning of the pandemic, because there was a collaborative spirit of, 'We're going to share information and share knowledge, and we're going to get to the answer,'" says Dybul, who participated in the Independent Panel for Pandemic Preparedness and Response. "Everyone's back in their corners. And that's very risky. Because everyone's just going back to the way we did it before, and that's not what we need to prepare for the next pandemic." ■

BIOLOGY

Mountain mice break altitude record for mammalian life

How rodents survive on summits is a mystery

By **Elizabeth Pennisi**

Three years ago on the border between Chile and Argentina, researchers saw a leaf-eared mouse (*Phyllotis vaccarum*) scurrying around the top of Llullaillaco, a 6700-meter volcano, where howling winds, below-freezing temperatures, a lack of plants, and scarce oxygen were thought to make life impossible for mammals. Now, the scientists have shown that the animal, common at lower altitudes, was not a fluke visitor to the peak.

The team has captured mice by the dozens on other Andean peaks, from the same species and others. Genomic comparisons of high-living *P. vaccarum* with their lower elevation relatives confirm that the former are permanent residents nearly 7000 meters above sea level, making them the highest dwelling mammal found so far. (The previous record holder was a single pika, a rabbit relative, spotted nearly 6200 meters up on Mount Everest a century ago.)

Andean leaf-eared mice (*Phyllotis vaccarum*) live above 6700 meters.

So far, the research team hasn't identified any genetic adaptations to high altitude in the mice. It's also unclear what they eat on the barren summits. (Lichen may be the answer, a preliminary analysis of stomach contents suggests.) Still, the evidence reported in *Current Biology* this week "lays to rest any doubt that mammals live at these really extreme altitudes," says Grant Mcclelland, a comparative physiologist at McMaster University who was not involved with the work. "It expands our understanding of the environmental limits of animals."

The cold temperatures and low oxygen at high altitudes have long been thought to rule out many resident animals, especially mammals. With their high energy demands, "Mammals in general are not very good at dealing with low oxygen environments," says Catherine Ivy, a comparative physiologist at the University of Western Ontario.

Yet 50 years ago, archaeologists studying Inca religious sites on Andes mountain summits came across mice naturally mummified by the dry cold. Researchers assumed the rodents had hitchhiked up with Inca people. But in 2013, two mountaineers summiting Llullaillaco filmed a mouse darting across the snow. So, starting in 2020, Jay Storz, an evolutionary biologist at the University of Nebraska-Lincoln, recruited Guillermo D'Elía, a systematist and mammalogist at the Austral University of Chile, and other South American colleagues for five expeditions up 21 peaks in the Central Andes. They found more mummified remains and trapped almost 500 live mice—half of them the leaf-eared—representing 18 species. For five species, the collected specimens established new elevation records, the team reported on 23 August on bioRxiv.

To confirm that the mice are high-altitude natives, Jeffrey Good, an evolutionary genomicist at the University of Montana, and colleagues sequenced genomes of *P. vaccarum* from different elevations. They found that the ones living above 6000 meters are close kin, with genomes more similar to one another than to their lower relatives, the team reports. "It helps solidify that these mice were actually breeding and residing at these altitudes," Ivy says.

The team also found other evidence, including several burrows above 6000 meters. And radiocarbon dating of the naturally mummified mice showed they lived no more than 350 years ago—long after the Incas stopped visiting the summit.

Hoping to understand how the high-altitude animals thrive so far up, Storz and colleagues have recently established lab colonies in Chile of mice from Llullaillaco and other locations surveyed. In the meantime, they are marveling at the unlikely new mountaintop habitats they have discovered. "Every new place we have a new mix of species," Storz says. "It's a little bit like the deepest depths of the ocean." ■

Richard Lichan Lekuterer's homestead, which on sat on the shore of Lake Baringo, is now submerged.

PHOTO: TONY KARUMBA/AFP VIA GETTY IMAGES

SCIENCE & THE LAW

In Kenya, rising lakes prompt novel climate change lawsuit

Residents around Lake Baringo say the government has failed to address flooding caused by a warming climate

By **Carey Baraka**

The town of Kampi Ya Samaki was once a bustling fishing and tourism center on the shores of Lake Baringo, one of a chain of major lakes nestled in the Great Rift Valley of western Kenya. Today, however, much of the town is submerged, with only the tops of houses, hotels, churches, and schools still visible. Seven islands used to sit just offshore. But the rising waters mean now "there are six," says Evans Limo, a local tour guide.

Such flooding has become a common sight throughout the Rift Valley, where lakes have swelled over the past decade. Lake Baringo, for example, has doubled in area since 2010 to about 26,000 hectares. Kenyan officials say the flooding has created "panic and anxiety" in lakeshore communities and affected nearly 400,000 people, with many forced to abandon their homes. In some communities, shifting shorelines have exposed residents to deadly attacks by crocodiles and hippos.

Now, a novel lawsuit brought by people living around Lake Baringo has put a spotlight on the question of whether climate change is to blame for the rising lakes—and whether Kenya's constitution and a landmark 2016 climate law obligate government agencies to compensate flooding victims. The goal of the court challenge, which was scheduled to get a hearing this week, is to "enforce the climate change duties of public officials," says Omondi Owino, the lead attorney representing the residents. The case, analysts say, is among the first in Africa to test a government's responsibility for helping its citizens cope with the impacts of climate change.

For years, residents and researchers were uncertain about the cause of the expanding lakes. Some believed the trend was tied to the complex geology and groundwater hydrology of the Great Rift Valley, a series of faults and ridges stretching 7000 kilometers from Lebanon across eastern Africa to Mozambique. Historically, they noted, the lakes have periodically shrunk and grown dramatically, to sizes even larger than seen today, perhaps because of crustal movements.

Others believed climate change was to blame. Average temperatures in eastern Africa have climbed by about 1°C over the past 50 years, and climate models have predicted that the region would become wetter as it warmed. But real-world data have complicated those predictions by showing that, although regional precipitation has ticked up in recent decades, droughts have also become more prevalent—a puzzle scientists have dubbed "the eastern Africa climate paradox."

To clarify the issues, in 2020 the Kenyan government commissioned a major inquiry into the causes and consequences of the flooding. Research teams visited the lakes and combed through data on land use, climate, and hydrology. In a 2021 report they concluded that a number of factors were likely contributing to lake expansion, including land use changes that have accelerated runoff and caused sediment to build up on lake bottoms. But, "The main reason for the rising water levels is climate change," the authors wrote. The report noted that precipitation had increased in upland areas, as the models predicted, especially in 2019, a particularly wet year. Rivers feeding the lakes had swelled, even as rainy, cooler weather reduced evaporation. "The inputs increased and the main output—which is evaporation—was reduced due to the cold spells and extended rains," says geologist Lydia Olaka of the Technical University of Kenya.

Mathew Herrnegger, a hydrologist of the University of Natural Resources and Life Sciences, Vienna, who was not involved the report, agrees with that analysis. Other studies, he notes, have concluded that "the Rift Valley lakes are very sensitive to changes in climate, which has led to complete desiccation but also high [water levels] in the past."

The 2021 report also suggested the government address the humanitarian crisis caused by the flooding, including by resettling displaced people and perhaps even buying submerged properties. But adequate help never materialized, alleges the lawsuit filed in 2022 by 66 members of the Ilchamus and Tugen communities against the county government of Baringo and national agencies.

That inaction, argue the residents and Kituo Cha Sheria, a Nairobi-based advocacy group, violates human rights provisions of Kenya's constitution and the 2016 climate law, which was the first such measure adopted by an African nation. Among other things, that law empowers the government to "mobilize … public and other financial resources for climate change response." The lawsuit alleges that government officials "failed, refused, or neglected" to "anticipate, prevent, or minimize" the impacts of climate change.

Government lawyers have rejected that claim in recent filings at the Environment and Land Court in Iten. Even if climate change is causing the flooding, they say the government is not responsible for paying damages because Kenya is not a major contributor to planetary warming. And the county government argues that it, too, is a victim of the flooding. It notes the rising lakes have swallowed numerous government facilities, such as police stations and schools, and reduced revenues from the tourism industry.

As *Science* went to press, each side was preparing for a 24 October court hearing. The outcome of the case is hard to predict, observers say, and any decision might not come for many more months. ■

Carey Baraka is a journalist in Nairobi, Kenya.

SCIENCE & SOCIETY

Israel-Hamas war shakes scientific community

Attacks have left many research laboratories without staff or in ruins

By **Michele Chabin**

Shock waves from the war between Israel and Hamas are rocking the scientific community in the region and around the world.

In Israel, universities are mourning scientists and others killed in the 7 October surprise attack by Hamas militants. Hamas, widely considered a terrorist organization, rules the Gaza Strip. Academic laboratories are emptying as foreign graduate students return home and young academics report for military service. Some science continues, says Asher Cohen, president of the Hebrew University of Jerusalem, but "we cannot maintain full research facilities."

Missiles from Israel's Iron Dome defense system intercept rockets launched from the Gaza Strip on 8 October.

In the Gaza Strip, researchers say many of the Palestinian enclave's few and already beleaguered scientific institutions have been damaged by Israeli air attacks, and scientists are struggling to find safe havens. "The priority isn't science—the priority is staying alive," says Marwan Awartani, president of the Palestine Academy for Science and Technology, which is based on the West Bank.

Elsewhere, the war is forcing research funders to remake plans. Last week, for example, the European Commission announced it would be giving all applicants an additional week or more to complete upcoming requests to several major grant programs. "We offer our support to those affected by the recent hostilities," Iliana Ivanova, European commissioner for innovation, research, culture, education, and youth, said in an 18 October statement.

Israel declared war on Hamas on 8 October, soon after the group's operatives murdered more than 1300 Israelis and foreign nationals, most of them civilians, and kidnapped 200 others. As *Science* went to press, the death toll in Gaza stood at more than 6000, according to Palestinian officials.

After Israel called up more than 300,000 reservists, colleges and universities in Israel postponed the start of the fall semester until at least early November. Some have turned their dormitories into housing for reservists or shelters for Israelis displaced by the war.

At Ben-Gurion University of the Negev, most of its 242 international postdoctoral students have departed, often at the insistence of their home countries, President Daniel Chamovitz said. The university is located just 41 kilometers from the Gaza border and within reach of rockets fired by Hamas and its allies. At least 50 Ben-Gurion students, staff, and faculty were killed in the Hamas attack, university officials say, including theoretical physicist Sergey Gredeskul and mathematician Viktoria Gredeskul, who were married. Others are being held hostage by Hamas.

The large number of dead has forced Chamovitz to develop a mourning protocol for the university. It ensures that at least one member of the administration attends every funeral and every shiva—the 7 days of Jewish mourning.

Several of Ben-Gurion's research labs are still functioning, but at less than peak capacity, Chamovitz said. Postdocs in data science or who are writing journal articles continue to work, and the university has decided to continue to pay scholarships for the next month, in the hope that the situation will stabilize.

At Hebrew University, biomedical researcher Yuval Dor says five of the 20 members of his laboratory, which focuses on diabetes and disease detection, have left for military service. "Everybody's distracted. … Everyone knows people who were hurt or killed or kidnapped," Dor says. "Experiments have pretty much come to a halt." The lab was also starting a new company, but Dor fears potential investors will now shy away.

Dor says he's been heartened by the "dozens and dozens" of messages of support he's received from colleagues abroad. Still, he says, "It is heartbreaking to shift from focusing on disease diagnosis to contemplating methods for identifying the DNA of victims."

At the Technion – Israel Institute of Technology, some international students have insisted on staying despite warnings to leave, says President Uri Sivan. Their support "is inspiring," he says.

In Gaza, the situation is dire, says Awartani, a former education minister for the Palestinian Authority, which presides over the West Bank. Even before the current war, Israel's blockade of Gaza made it extremely difficult for Palestinian researchers to receive funding or attend conferences overseas, and for foreign researchers to work in Gaza's universities. Israel often blocked the importation of scientific equipment, citing security concerns.

Now, much of Gaza's limited scientific infrastructure is in ruins, Awartani says. Two major institutions—the Islamic University of Gaza and Al-Azhar University—have suffered extensive damage from Israeli air strikes. "It is clear that many students and faculty are dead or wounded," Awartani says.

"The long-term ramifications of this war on Gaza's higher education system are incalculable," the Palestine Academy for Science and Technology said in a statement. "The psychological, intellectual, and cultural fabric necessary for a thriving academic community has been torn asunder, and it will take years, if not generations, to mend." ■

Michele Chabin is a journalist in Israel.

FEATURES

HIGH HOPES FOR SHORT CORN

Plants bred or engineered to be short can stand up better to windstorms. They could also boost yields and benefit the environment *By* **Erik Stokstad**

On 10 August 2020, a record-breaking windstorm raced across Iowa, the big buckle of the U.S. Corn Belt. Gusts up to 225 kilometers per hour flattened fields and buildings, with losses totaling an estimated $12 billion across several states. "It was devastating," says Kelly Gillespie, a crop physiologist with Bayer Crop Science, who recalls driving by ruined houses and wrecked silos. About 16% of Iowa's corn and soybean crops were damaged or destroyed. "Corn snapped and broken for as far you could see."

In the midst of the destruction, though, Gillespie saw glimmers of hope. When she and other researchers visited fields of experimental corn plants developed by Bayer, they saw that most of the plants had resisted the force of the storm and remained standing. The source of their strength? They were shorter.

To an interstate traveler—or anyone lost in a corn maze—the most impressive feature of corn is its stature. Modern corn can grow twice as tall as a person, but height has drawbacks, making the plants vulnerable to wind and more difficult for farmers to tend. Plant scientists think corn can be improved by making it shorter, and leading seed companies are doing that through both conventional breeding and genetic engineering. Bayer has launched a short variety in Mexico, another company is selling its versions in the United States, and more are getting involved.

Researchers say short corn could be a boon for farmers. "This is an idea long overdue," says Rex Bernardo, a corn breeder and geneticist at the University of Minnesota. "I'm quite excited." In addition to bolstering wind resistance, the short stature will allow farmers to drive tractors into their fields longer into the summer to add late-season fertilizer or fungicides to boost the harvest. And according to Stine Seed, a company that has pioneered short corn, the new crop can increase yield even further because it can be planted more densely.

Bayer is now wrapping up its final round of U.S. trials, which took place this summer across 12,000 hectares in Iowa and three nearby states. The company plans to start selling seed to U.S. farmers next year,

Farmers can drive a tractor through fields of corn bred to be shorter, allowing them to add fertilizer later in the year.

and at least one other large seed company, Corteva Agriscience, will soon follow suit with its own varieties. Bayer's corn was conventionally bred with the Midwest Corn Belt in mind, but the company has recently developed a transgenic version it can quickly and efficiently tailor to other regions. In June, it was approved by the U.S. Department of Agriculture (USDA), a key regulatory step.

The new corn varieties join a long tradition of improving crops by shrinking them. In the 1950s and '60s, plant breeders created semi-dwarf varieties of rice and wheat, allowing plants to bear more grain without collapsing. These high-yielding varieties were key to the Green Revolution, and they prevented famines in developing world. Now that researchers have pieced together the complex interplay of genes and hormones that control corn growth, the time may be ripe for corn, too, to go big by getting shorter.

CORN (*ZEA MAYS*) was a much different plant when it was domesticated in southern Mexico some 9000 years ago. Compared with modern corn, its closest living relative, a lanky wild grass called teosinte, looks disheveled, with multiple stalks and tassels that mature into many small cobs with just two rows of seeds. In what would be a nightmare for farmers, teosinte's seeds fall off, rather than clinging to the cob for easy harvest. Some of the oldest corn cobs uncovered by archaeologists are also tiny and fragile. But after millennia of selective breeding—first by Indigenous farmers, and later in universities and multinational companies—corn has become one of the world's dominant food crops. In the United States, it brings in about one-third of income from all crops, earning farmers $89 billion in 2022.

With success came stature. Over the past half-century, as breeders selected for larger ears with more numerous and plumper, heavier kernels, they ended up choosing big plants with many leaves for photosynthesis and tall stalks. Modern corn plants can grow up to 4 meters high, although the dominant starchy varieties in the U.S.—88% is used for livestock and ethanol—range between 2.5 and 3.5 meters.

Tall plants have an inherent weakness, however. The weight of the ears, usually more than halfway up the stalk, makes the plant vulnerable to snapping or tipping over in strong winds. Making corn shorter isn't the only remedy, notes Tony Vyn, an agronomist at Purdue University. Breeders have also selected for stiffer stalks, as well as for resistance to stalk rot, which made earlier corn varieties more prone to snap under the weight of their grain or blow over in windstorms.

Even so, between 2001 and 2016, about 800,000 hectares of corn fields were damaged by high winds, according to U.S. government crop insurance claims. That's not much compared with the 38 million hectares damaged by drought during the same period, but enough that companies see it as a selling point for short corn.

Bayer has used conventional breeding to create three short hybrid varieties that were tested for the U.S. market this summer and are now being harvested by about 300 farmers. Field trials demonstrated better wind resistance compared with tall corn, Gillespie says. "Seeing the standing corn for the first time just gave me goosebumps." The 2020 windstorm provided an even more strenuous test. Afterward, Gillespie and colleagues analyzed 14 experimental fields that were hit by heavy winds; on average 25% of short corn was damaged compared with 50% of tall corn, they reported in *Crop Science* in August 2022.

Corteva, which owns the famous seed company Pioneer Hi-Bred International, is aiming for similar results with its conventionally bred short corn, which it hopes will be ready for market in a few years. Corteva researchers test the strength of new hybrids by exposing them to gusts of up to 160 kilometers per hour from a wind machine.

Breeding shortness into an existing variety of corn typically takes 5 years or more. To shortcut this process, Bayer has turned to genetic engineering. Speed is important because Bayer sells about 200 varieties of hybrid corn tailored to various conditions, such as climate and disease threats. Creating short versions of all those varieties through conventional breeding, and making enough hybrid seed to sell, would be prohibitively time consuming.

Bayer's genetic engineers have focused on a plant growth hormone called gibberellin. At first the approach looked unpromising. Decades ago researchers knocked out the genes that control early steps in gibberellin synthesis, but the result was plants that had deformed flowers and abnormal ears. Since then, scientists have adapted an approach previously used to engineer shorter versions of tomato, apple, banana, and other crops. They figured out how to suppress corn genes that code for enzymes known as gibberellin 20 oxidases, which put the biochemical finishing touches on the hormone in particular parts of the plant. This way, the levels of gibberellin can be lowered in the stalks, but not the flowers, stunting the plants' growth without undermining their productivity.

Initial attempts resulted in corn plants that were far too short to be commercially viable. But after considerable tinkering, Gillespie and her colleagues solved the problem in collaboration with the company BASF. They added DNA that encodes microRNAs, small molecules that can suppress genes. In this case the targets were two genes that regulate the creation of gibberellin mainly in the leaves and stalk, resulting in plants that were one-third shorter. To avoid shrinking the leaves as well, the researchers delivered the mutant alleles with a genetic switch from a rice virus that is most active in the stalk.

Field studies in 2019 and 2020 at two locations in Illinois confirmed the manipula-

tion did not affect the ears; they weighed the same as those from typical tall varieties, Gillespie and colleagues reported in March 2022 in the *Plant Biotechnology Journal*. This successful change in stature by interfering with gibberellin was "the first of its kind in corn," says Guri Johal, a corn geneticist at Purdue who has discovered a mutation, called D16, with a similar effect.

Some of the time saved with transgenic corn is lost because of stricter regulatory requirements. Companies must provide field and laboratory data to USDA, which assesses whether genetically modified crops pose a threat to agriculture, and they consult with the U.S. Food and Drug Administration for checks on food safety. Because Bayer's version will be paired with a transgenic trait for pest resistance (Corteva's likely will, too), the Environmental Protection Agency will have to evaluate it as well. The reviews can take years.

To speed up the process, Bayer and other companies are working on short corn made with gene editing, an approach that can alter gibberellin without adding genes from other species. Corn varieties created this way will face fewer regulatory hurdles in the U.S. and some other countries, but they're still years away.

IN SINALOA, MEXICO, where luxuriant corn fields carpet the plains between the Sierra Madre mountains and the Gulf of California, short corn is already having an impact. The farms are some of the most productive in Mexico but have long suffered from large, persistent windstorms called derechos. In 2007, Bayer began to breed a short corn for the region's farmers by selecting varieties with a mutation that limits transport of another growth hormone, auxin, in the stalk. The resulting plants are about one-third shorter than normal, and less likely to fall over or snap in a derecho. (This variety, which went on the market in 2020 as VITALA, isn't suitable for use in the United States because it is tailored to the soil and climate in Sinaloa, as well as the day length at that latitude.)

At a February agricultural trade show in Culiacán, Sinaloa's capital, Bayer highlighted its short corn in large party tents surrounded by corn plants. Salesmen in black jeans and white company shirts pitched VITALA as a package, combined with agronomic advice and digital maps based on satellite imagery and data streams from the farmers' high-tech tractors and combines. The combination of technology and improved seed can push yields to 16 tons of corn per hectare, Bayer says, compared with the 12 tons typical for Sinaloa. Bayer expects about 6000 hectares of VITALA short corn to be planted in Sinaloa next year.

Whether conventionally bred or genetically engineered, short corn has other advantages, says Fred Below, a crop physiologist at the University of Illinois Urbana-Champaign, who has received funding from Bayer to study its new hybrids. Because the plant puts fewer resources into its stalk, it can divert more into roots—though Vyn notes that the extent to which this actually happens hasn't been fully studied. A more developed root system would help plants better tolerate drought and extract fertilizer from the soil.

In U.S. field trials, the short corn benefited from nitrogen applied later in the season when it would be harder to add to tall corn, Vyn, Below and colleagues reported in *Field Crops Research* in May. That could have environmental benefits. USDA scientists have reported that water quality near farm fields can improve 15% to 20% when some of the fertilizer is applied later in the season. That's when corn plants are enlarging their kernels and need it most, reducing the excess that can leach into groundwater or run off into streams. Excess nitrogen pollutes groundwater and causes algal blooms in lakes—even as far away from the Corn Belt as the Gulf of Mexico. Making better use of nitrogen fertilizer may turn out to be the biggest benefit of short corn, says Kendall Lamkey, a corn breeder at Iowa State University.

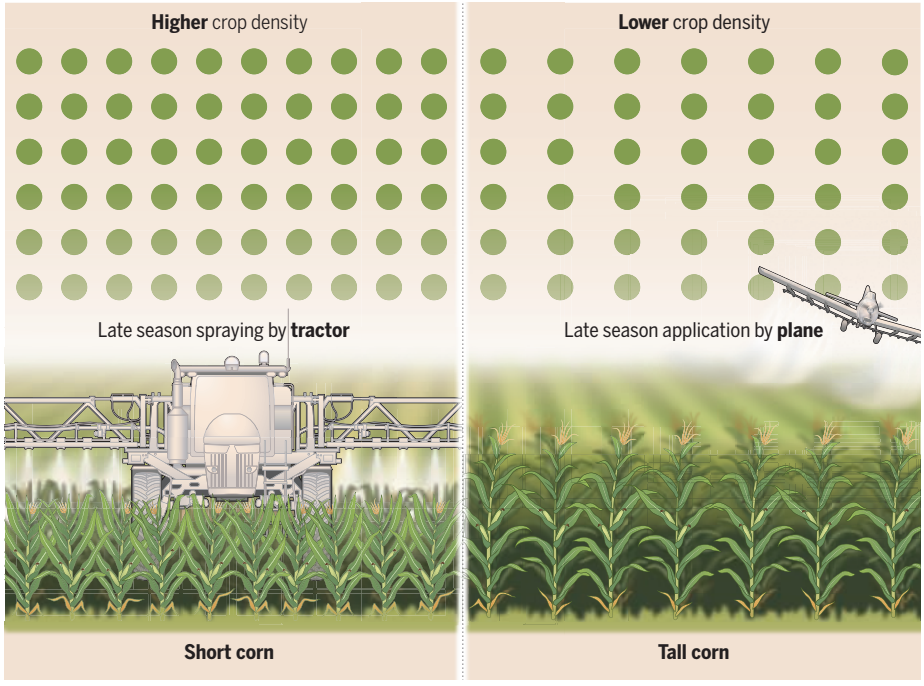
Additional environmental benefits could accrue if short stature allows farmers to spray fungicides on mature corn plants from tractors rather than airplanes, which can lead to greater drift away from the fields, potentially exposing people and wildlife. Although downsides may yet materialize, Below says he's optimistic about the potential: "I'm convinced it's going to be more advantages than disadvantages."

FOR U.S. FARMERS, the prospect of greater yield may be the biggest lure. Stine Seed, based in Iowa, already sells short corn in the United States. Its varieties are not planted on a large scale, in part because Stine is a relatively small producer of corn seed. But Stine says its hybrids can exceed the yield of tall corn by up to 10% in optimal conditions.

Founder Harry Stine grew up on a small Iowa farm and started breeding corn and soybeans in the 1970s. Larger companies like Monsanto licensed the improvements he made in soybeans, making Harry Stine a fortune; now in his early 80s, he is Iowa's only billionaire.

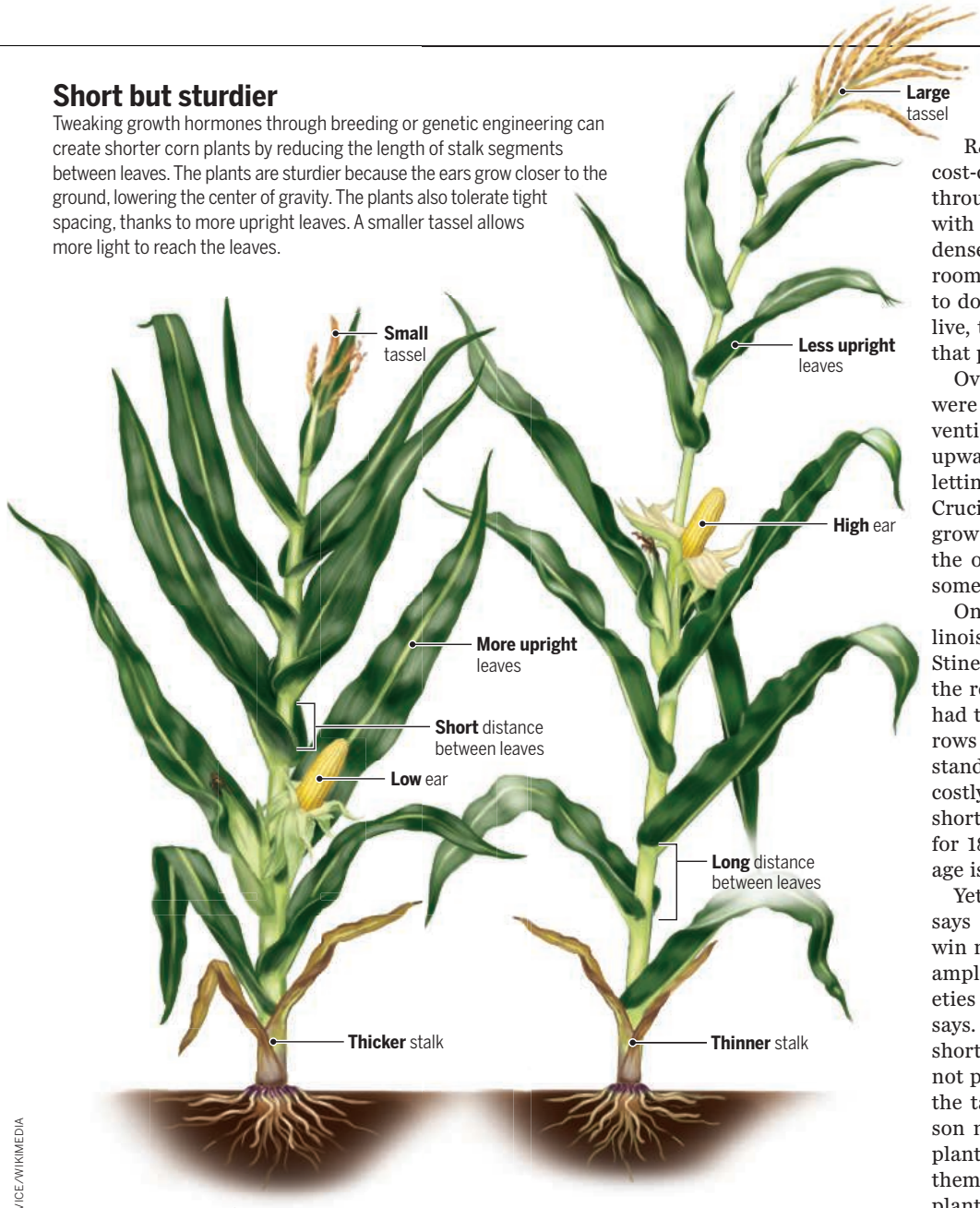
The advantages of being short

Short corn can be planted more densely, increasing yields per hectare (aerial view, top), although the trade-off is greater risk from dry weather. Short plants also resist strong winds better. Regardless of density, short plants enable farmers to apply fertilizer or fungicide in late summer by tractor rather than by plane (bottom), potentially reducing environmental impacts.



Short but sturdier

Tweaking growth hormones through breeding or genetic engineering can create shorter corn plants by reducing the length of stalk segments between leaves. The plants are sturdier because the ears grow closer to the ground, lowering the center of gravity. The plants also tolerate tight spacing, thanks to more upright leaves. A smaller tassel allows more light to reach the leaves.



Severe windstorms wreak havoc on fields of tall corn, like this one in Iowa laid low by a derecho in 2020.

In the company's early days, however, Stine had far less to spend on R&D than the big seed companies. That cost-cutting led to an unintended breakthrough with corn. The standard practice with breeding trials was to plant corn densely, then thin the plots to give plants room to grow. Stine didn't have the labor to do this weeding, so he let all the plants live, then selected the corn from the plots that performed the best.

Over the years, this led to hybrids that were about one-third shorter than conventional corn. Their leaves are angled upward more, and the tassels are smaller, letting more light reach the lower leaves. Crucially, this means more plants can be grown per hectare, potentially increasing the overall harvest. "I think Harry's onto something," Lamkey says.

On his family farm in north-central Illinois, Darron Judd has been growing Stine's short corn for several years. To get the recommended density in his fields, he had to modify his tractor to plant seeds in rows 38 centimeters apart rather than the standard 76 centimeters. The change was costly, but Judd likes the sturdiness of the short corn. "It just stands great." He aims for 18 tons per hectare. The Illinois average is 13 tons.

Yet Stine Seed President Myron Stine says several factors have made it hard to win new customers for short corn. For example, farmers like to compare new varieties to what they're used to growing, he says. If they do this by alternating rows of short and tall corn, the shorter hybrid will not perform as well because it's shaded by the taller corn. Lamkey says another reason many farmers have been reluctant to plant corn more densely is that it forces them to bet on the weather: in a dry year, plants packed into a rain-fed field can literally run out of water. And Seth Murray, a corn geneticist and breeder at Texas A&M University, notes that farmers have to buy more seed to plant densely—offsetting some of the profit from higher yield.

For Bayer and Corteva, those hurdles may prove less daunting. Those two companies account for 70% of the U.S. corn seed market, and they have more resources to promote proper planting of short corn.

In Iowa, at least, they expect to find receptive ground. Farmers there have become "hyperfocused" on the risk of derechos, Lamkey says. Those who turn to short corn to reduce the risk of loss may reap other benefits, in yield and environmental protection.

Decades after wheat and rice were transformed in the Green Revolution, corn is getting its turn. ■

CREDITS: (GRAPHIC) N. BURGESS/*SCIENCE*; (PHOTO) LISA SCHMITZ/NATIONAL WEATHER SERVICE/WIKIMEDIA

EVOLUTION

Menopause in chimpanzees

Signs of menopause in wild chimpanzees provide insights into human evolution

By **Michael Cant**

Humans are one of only a handful of mammals in which females cease reproducing long before they die. Explaining how menopause evolved is a challenge because it is not obvious why selection should favor genes that extend life span past the end of reproduction. It is also unclear why menopause should have evolved in certain toothed whales and in humans but not other primates. On page 416 of this issue, Wood *et al.* (*1*) report that female chimpanzees (*Pan troglodytes*) living in a relatively undisturbed population in Uganda show physiological signs of menopause and survive long past the end of reproduction. This observation means that the role of grandmothers in the evolution of menopause may need to be reconsidered.

Studies of long-lived animals have revealed substantial differences in behavior between different breeding populations. For example, some killer whale (*Orca orcinus*) populations feed exclusively on fish, whereas other populations prey mainly on mammals. Chimpanzees in some populations use stone tools to crack nuts, and those in other populations use wooden tools for the same purpose. This interpopulation variation provides the raw material to test theories about how these differences evolve. However, measuring the pattern of survival and reproduction of long-lived animals takes decades of work,

Female chimpanzees (*Pan troglodytes*) from the Ngogo community in Uganda can live for decades after losing the ability to reproduce.

so information about a species' life history is usually based on one or (at most) very few populations. New data from less well-known populations have the power to upend assumptions about the life history of even the most familiar species.

Women living in natural fertility populations (populations without access to modern medicine or technology) often live many years of active life after the loss of reproductive ability, even in the harshest ecological conditions. Among the Hadza hunter-gatherers of northern Tanzania, for example, women have their last baby at an average of 41 years of age, but women who reach this age can expect to live into their late 60s (*2*). Thus far, a comparable period of postreproductive life has been observed in only five other mammals in the wild—killer whales, short-finned pilot whales (*Globicephala macrorhynchus*), narwhals (*Monodon monocerus*), beluga whales (*Delphinapterus leucas*), and false killer whales (*Pseudorcus crassidens*) (*3*).

The main explanation for how these exceptionally long postreproductive life spans evolved is based on the theory of kin selection, which is the idea that organisms can propagate copies of their genes not just by reproducing themselves but also by helping their genetic kin to reproduce (*4*, *5*). Older females could evolve to cease breeding before the end of life if by doing so they could increase their children's reproductive success (or "fitness"). For example, by entering a postreproductive state, older females could free up food and other resources that are required for reproduction (*6*) or care for their grandchildren directly (*7*). Studies of killer whales and some human populations have found evidence for both types of fitness benefit (*8*).

Kin selection has also been used to explain why chimpanzees do not experience menopause. Fertility in female chimpanzees ends around 50 years of age, but in the best-studied wild populations, females rarely reach this age (*9*). This lack of a postreproductive life span fits the kin selection theory because unlike humans and killer whales, older chimpanzee females are not hard-working grandmothers. Moreover, because young female chimpanzees disperse from their natal community at adolescence while the males remain at home, there are usually no breeding daughters around for chimpanzee grandmothers to help. Grandmothers could help their sons to reproduce, but they do not appear to.

The work of Wood *et al.* challenges the kin selection argument. They report that unlike other chimpanzee populations, female chimpanzees in the Ngogo community of Uganda's Kibale National Park frequently survive long past the age of reproduction. Females rarely gave birth after the age of 45 years but often survived into their 50s or 60s. The transition to this postreproductive state was marked by increasing urine levels of follicle-stimulating hormone and luteinizing hormone and decreasing urine levels of estrogens and progestins—the same endocrinological changes that are associated with human menopause.

Wood *et al.* calculated a metric called postreproductive representation (*PrR*), which is the average proportion of the adult life span that is spent in a postreproductive state. For most mammals, including other chimpanzee populations, *PrR* is close to zero (0.001 to 0.04) (*1*, *3*). But in the Ngogo population, Wood *et al.* found a *PrR* of 0.2, meaning that females on average live 20% of their adult years in a postreproductive state. This value is close to that observed in humans (0.3 to 0.47) and some species of whales (0.24 to 0.4).

Wood *et al.* consider two possible explanations for their findings. Chimpanzees and other mammals can have extended postreproductive life spans in captivity where they are protected from natural predators and pathogens (*10*). Therefore, if the Ngogo chimpanzees are similarly protected, they might also live artificially long lives. For example, predation by leopards or competition from other chimpanzee groups may have been more intense in the past (*1*). In this case, the high *PrR* value in the Ngogo females would reflect the relative inflexibility of the fertility schedule to changes in ecological conditions compared with life span. Such inflexibility is observed in humans, in which the average age at menopause has remained relatively stable (at around 50 years) for more than a century (*11*), despite major improvements in nutrition, health care, and expected life span.

Alternatively, the Ngogo chimpanzees might be living in conditions that are more representative of their evolutionary past compared with other populations. The Ngogo population is relatively remote and undisturbed by human activities, such as logging and hunting. Ngogo chimpanzees are also less exposed to potentially devastating human diseases than other chimpanzee populations. Wood *et al.* argue that previous research may have underestimated the life span of wild chimpanzees before the start of modern human influence.

If chimpanzees evolved to undergo menopause, theories to explain it may need to focus on the similarities between ape and toothed-whale societies. One of the consequences of the female-biased dispersal pattern in chimpanzees is that adult females become more closely related to others in their group as they get older. The same is true in killer whales and short-finned pilot whales, in which neither sex disperses but mating occurs between groups (*12*). Female-biased dispersal might also have been common in ancestral human populations, although this is still debated. Given this pattern of kinship dynamics (*12*), theoretical models predict that older females may evolve to cease reproduction to avoid competition with younger females. This type of fitness benefit has been detected in killer whales (*13*) and in some human populations (*14*) but not in others (*2*). Wood *et al.* suggest that selection to avoid reproductive competition between generations may also have shaped the fertility schedule of chimpanzees.

Older postreproductive chimpanzees might also provide fitness benefits to their local group through their experience and knowledge. In killer whales, postreproductive females lead the group in search of food, particularly when prey is scarce (*15*). Could older female chimpanzees confer similar benefits on others in their group? Further study of the postreproductive Ngogo females would help answer this question.

The study of Wood *et al.* both illuminates and raises questions about the evolution of menopause. It also highlights the power of difficult long-term field studies—often run on small budgets and at constant risk of closure—to transform fundamental understanding of human biology and behavior. ■

Centre for Ecology and Conservation, University of Exeter, Penryn, UK. Email: m.a.cant@exeter.ac.uk

REFERENCES AND NOTES

1. B. M. Wood *et al.*, *Science* **382**, eadd5473 (2023).
2. N. B. Jones, *Demography and Evolutionary Ecology of Hadza Hunter-Gatherers*, vol. 71 of Cambridge Studies in Biological and Evolutionary Anthropology series (Cambridge Univ. Press, 2016).
3. S. Ellis *et al.*, *Ecol. Evol.* **8**, 2482 (2018).
4. K. Hawkes, *Am. J. Hum. Biol.* **15**, 380 (2003).
5. W. D. Hamilton, *J. Theor. Biol.* **12**, 12 (1966).
6. M. A. Cant, R. A. Johnstone, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5332 (2008).
7. P. S. Kim, J. S. McQueen, K. Hawkes, *J. Theor. Biol.* **461**, 84 (2019).
8. D. P. Croft, L. J. N. Brent, D. W. Franks, M. A. Cant, *Trends Ecol. Evol.* **30**, 407 (2015).
9. M. E. Thompson *et al.*, *Curr. Biol.* **17**, 2150 (2007).
10. J. G. Herndon *et al.*, *Age* **34**, 1145 (2012).
11. D. J. Frommer, *BMJ* **2**, 349 (1964).
12. R. A. Johnstone, M. A. Cant, *Proc. Biol. Sci.* **277**, 3765 (2010).
13. D. P. Croft *et al.*, *Curr. Biol.* **27**, 298 (2017).
14. M. Lahdenperä, D. O. S. Gillespie, V. Lummaa, A. F. Russell, *Ecol. Lett.* **15**, 1283 (2012).
15. L. J. N. Brent *et al.*, *Curr. Biol.* **25**, 746 (2015).

10.1126/science.adk7119

MOLECULAR BIOLOGY

Genome editing with retroelements

RNA-guided DNA writing enzymes offer promise for programmable gene insertion

By **Stephen Tang** and **Samuel H. Sternberg**

Genome editing methods generally involve two critical steps: targeting a DNA sequence of interest and introducing the desired genetic modification. Capabilities for programmable DNA targeting have advanced rapidly over the past decade, owing primarily to the retooling of bacterial RNA-guided CRISPR-Cas systems. However, efficient and precise modification of the target site remains challenging, particularly for large-scale perturbations such as the insertion of gene-sized DNA payloads. To address this limitation, an emerging body of research is now drawing inspiration from mobile genetic elements (MGEs), DNA segments that have exploited diverse targeting and integration strategies to propagate in the genomes of their hosts for billions of years. Of particular interest are retroelements, which mobilize using an RNA intermediate. Recent investigations into the structure and function of retroelement-encoded protein-RNA complexes have resolved key molecular details of their targeting and integration mechanisms (*1, 2*), unveiling exciting opportunities to reengineer them for programmable DNA insertion.

MGEs are present in all organisms and encode enzymes that facilitate the movement of their own DNA within and between genomes. Two major categories of MGEs—DNA transposons and retroelements—can be distinguished by their spreading mechanisms. DNA transposons typically encode transposase enzymes that mobilize DNA by excising sequences from their original location and integrating them into a target locus ("cut-and-paste"). By contrast, retroelements rely on the host cell's RNA polymerase to generate an RNA copy of the element, which in turn serves as a template for complementary DNA (cDNA) synthesis by a retroelement-encoded reverse transcriptase (RT). This cDNA product is then incorporated into the genome through a range of mechanisms, creating a new copy of the original DNA segment ("copy-and-paste").

MGEs are natural candidates for genome editing applications. Their ubiquitous presence reflects a fine-tuned capacity for safe and efficient DNA integration into host genomes. However, MGEs generally exhibit either random or fixed specificity for integration sites, which limits their utility for programmable DNA insertion. This contrasts with CRISPR-based genome editing, in which the Cas9 nuclease can be directed by a user-defined guide RNA (gRNA) to target almost any DNA sequence of interest. But, in conventional genome editing with Cas9, DNA targeting and editing constitute separate processes: Cas9 recognizes and cleaves the target sequence to generate a DNA double-strand break (DSB), activating cellular DNA repair pathways that incorporate permanent sequence edits. This process introduces substantial heterogeneity and can lead to unwanted by-products, such as large chromosomal deletions and translocations, resulting in safety concerns around DSB-generating genome editing methods (*3*). CRISPR-Cas systems and MGEs thus exhibit complementary strengths and weaknesses with regard to target site specificity and genotoxicity. The ideal genome engineering tool would merge their capabilities, directly coupling programmable DNA targeting with DSB-free modification of the target locus.

Indeed, approaches that combine CRISPR-mediated DNA targeting with transposase-mediated DNA insertion have emerged in the past 5 years (*4–6*). These methods use various strategies to recruit transposase enzymes to gRNA-specified target sites, where they catalyze the integration of a donor DNA molecule supplied on a plasmid. Targeted DNA insertion with CRISPR-associated transposases (CASTs) represents a promising area of investigation, with demonstrated success for DNA cargoes up to 15 kilobases (*6*). Analogous retroelement-mediated approaches, which could encode the donor molecule on RNA rather than DNA, might offer additional advantages for clinical applications. Whereas DNA cargoes for genome editing are often delivered to target tissues by adeno-associated viruses (AAVs) or other vectors, RNA cargoes can be delivered by encapsulation in lipid nanoparticles. Lipid nanoparticles are less immunogenic than AAVs, and they present less risk for off-target editing owing to the transient nature of their RNA cargoes (*7*), which are degraded by the recipient cell. Thus, retroelement-based genome editing tools, which do not require any DNA components and can be packaged in lipid nanoparticles, offer a distinct safety advantage over tools that require an exogenously provided DNA donor molecule.

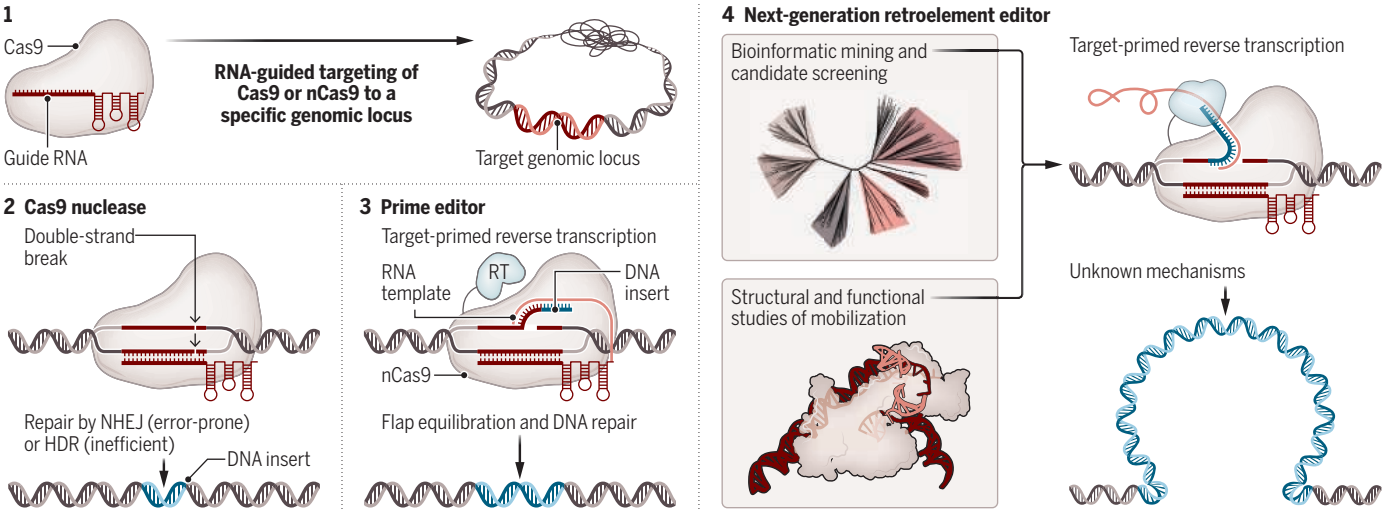
Retroelements have been a part of the genome engineering toolkit for over two decades, though until recently they have demonstrated limited utility for targeted DNA insertion. For example, targetrons are modified bacterial group II introns that home to user-specified DNA sequences through base-pairing interactions with the intron RNA, which catalyzes reverse splicing into the target site (*8*). Although targetrons have been used for targeted gene disruption, poor tolerance of synthetic cargoes has hindered efforts to engineer them for transgene insertion. Retrons, a more recent addition to the toolkit, are bacterial retroelements that have been engineered to produce donor DNAs from RNA templates for homology-directed repair of CRISPR-Cas9 cleavage sites (*9*). However, retron-mediated approaches have only been applied to small-scale edits, and their reliance on the generation of a DSB poses safety risks for clinical gene editing. Targetrons and retrons have thus demonstrated the promise of harnessing retroelements for biotechnology, but they fall short of the ultimate capability for a retroelement-mediated genome editing tool: to directly write any new DNA sequence into any genomic target site.

The development of prime editing represents the most important breakthrough to date toward achieving this capability. In prime editing, a nickase variant of Cas9 (nCas9) is directed by a prime editing guide RNA (pegRNA) to cleave a single DNA strand at the genomic target site, exposing a free DNA end from which an RT enzyme polymerizes a new DNA sequence using the pegRNA as a template. Cellular repair factors incorporate the new DNA strand into the genomic target site, resulting in permanent integration of the pegRNA-specified sequence (*10*). Prime editing avoids the generation of DSBs and can install all types of point mutations, small insertions, and small deletions. Nonetheless, editing efficiencies remain low for many applications, and the template size that can be accommodated is limited to ~40 base pairs or less. Recent advances have addressed this size limitation by employing multiple pegRNAs or combining prime editing with site-specific recombinases (*11*)—another class of enzymes used by MGEs to perform DNA integration—but at the cost of adding complexity and reaction steps to the

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA.
Email: shsternberg@gmail.com

A path to retroelement-based programmable gene insertion

Genome editing tools commonly use Cas9 directed by a guide RNA to recognize the target site (**1**) but employ distinct strategies for the DNA editing step. In conventional nuclease-mediated editing (**2**), a DNA double-strand break is made by Cas9 and repaired by nonhomologous end joining (NHEJ) or homology-directed repair (HDR). In prime editing (**3**), nickase Cas9 (nCas9) generates a free DNA end from which a reverse transcriptase (RT) writes a new RNA-templated DNA sequence. Further exploration of retroelements could enable RNA-templated integration of larger DNA payloads (**4**).



overall editing pathway.

Notably, the mechanism of DNA writing utilized by prime editors, known as target-primed reverse transcription (TPRT), is the same mechanism used to mobilize entire genes by a family of retroelements that is widespread in nature (*12*). Harnessing these elements, known as non–long terminal repeat (LTR) retrotransposons, has the potential to substantially advance gene insertion technology. But doing so will require a detailed understanding of their native mobility mechanisms, including how they select their respective RNA substrates and genomic targets and how they write their cargoes into target sites. Fortunately, the mobilization of non-LTR retrotransposons is relatively well studied owing to their pervasiveness—for example, the long interspersed nuclear element 1 (LINE-1) retrotransposon makes up nearly 20% of the human genome (*13*). Intriguingly, whereas LINE-1 undergoes largely untargeted mobilization, the R2 non-LTR retrotransposon, which is abundant in arthropods, exhibits stringent target site preference for the 28*S* ribosomal RNA locus (*14*). But how the R2 protein assembles with R2 RNA to initiate TPRT at this site, and whether R2 can perform TPRT with a synthetic RNA template, was unclear until earlier this year.

Two recent studies used cryo–electron microscopy to determine the three-dimensional structure of the R2 RT from *Bombyx mori* (domestic silk moth) bound to R2 RNA and the 28*S* DNA target (*1, 2*). This revealed specific molecular interactions in the RT-RNA-DNA complex, indicating that the retrotransposon selects its RNA template and DNA target through sequence-dependent recognition by the R2 protein. These structural details were used to engineer a synthetic R2 RNA that included the minimal recognition sequence and a "cargo" RNA appendage. The R2 RT recognized this engineered substrate and could be redirected by Cas9 to incorporate the cargo at non-28*S* target sequences in biochemical reactions (*1*).

A strong foundation has been laid for leveraging retroelements as targeted gene insertion tools. However, important gaps in knowledge remain. For example, genomic integration of cDNAs generated by the R2 RT requires the synthesis of a second cDNA strand (*14*), but the mechanistic details of this process are unclear. Additional studies aimed at resolving the molecular requirements for the final steps of R2 retrotransposition will be crucial for engineering R2 to perform transgene insertion in cells. Furthermore, systematic approaches for profiling the accuracy of TPRT have not been reported. The development of such methods will be critical for evaluating the safety of retroelement-mediated genome editing strategies.

A vast repertoire of retroelements exists in nature, yet only a few have been investigated for their genome editing potential. Certain RT enzymes with as-yet-unknown properties might be inherently well-suited for genome editing tasks. Systematic mining of retroelement diversity through bioinformatic analyses and pooled screening will be invaluable for the identification of RT enzymes with desirable characteristics (see the figure). This strategy is already bearing fruit for the development of compact prime editors (*15*). It is also worth noting that retroelements have already played an outsized role in shaping the human genome, roughly half of which is derived from MGEs (*13*). Thus, in developing retroelement-based DNA editing tools, humans are repurposing nature's original genome engineers. ■

REFERENCES AND NOTES

1. M. E. Wilkinson, C. J. Frangieh, R. K. Macrae, F. Zhang, *Science* **380**, 301 (2023).
2. P. Deng *et al.*, *Cell* **186**, 2865 (2023).
3. J. Y. Wang, J. A. Doudna, *Science* **379**, eadd8643 (2023).
4. J. Strecker *et al.*, *Science* **365**, 48 (2019).
5. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, *Nature* **571**, 219 (2019).
6. G. D. Lampe *et al.*, *Nat. Biotechnol.* 10.1038/s41587-023-01748-1 (2023).
7. A. Raguram, S. Banskota, D. R. Liu, *Cell* **185**, 2806 (2022).
8. H. Guo *et al.*, *Science* **289**, 452 (2000).
9. E. Sharon *et al.*, *Cell* **175**, 544 (2018).
10. A. V. Anzalone *et al.*, *Nature* **576**, 149 (2019).
11. P. J. Chen, D. R. Liu, *Nat. Rev. Genet.* **24**, 161 (2023).
12. D. D. Luan, M. H. Korman, J. L. Jakubczak, T. H. Eickbush, *Cell* **72**, 595 (1993).
13. L. M. Payer, K. H. Burns, *Nat. Rev. Genet.* **20**, 760 (2019).
14. T. H. Eickbush, D. G. Eickbush, *Microbiol. Spectr.* **3**, 3.2.02 (2015).
15. J. L. Doman *et al.*, *Cell* **186**, 3983 (2023).

ACKNOWLEDGMENTS

The authors thank D. Zhang and T. Wiegand for assistance with the figure, as well as A. Chen for insightful comments on the manuscript. S.T. is supported by a Medical Scientist Training Program grant (5T32GM145440-02) from the National Institutes of Health (NIH). S.H.S. is supported by NIH grant DP2HG011650, a Pew Biomedical Scholarship, a Sloan Research Fellowship, an Irma T. Hirschl Career Scientist Award, and a generous startup package from the Columbia University Irving Medical Center Dean's Office and the Vagelos Precision Medicine Fund. S.H.S. is an inventor on patents and patent applications related to CRISPR–Cas systems and uses thereof, a cofounder of and scientific adviser to Dahlia Biosciences, a scientific adviser to Prime Medicine and CrisprBits, and an equity holder in Dahlia Biosciences and CrisprBits.

10.1126/science.adi3183

GRAPHIC: A. MASTIN/*SCIENCE*

NEUROSCIENCE

Encoding identity in the marmoset

Hippocampal cells integrate multisensory input to represent the identity of others

By **Sylvia Wirth**

Humans identify other individuals using traits including their face, voice, or gait. By combining these perceptual cues with semantic information, such as the person's name and memories of previous encounters with that person, humans form a conceptual representation of the other's identity. Neurons in the hippocampus—a brain region involved in memory—can bind this multimodal information into a distinct conceptual representation of another's identity by responding equally to different views of a person's face and reading or hearing their name (*1*). On page 417 of this issue, Tyree *et al.* (*2*) report that such multisensory integration is not specific to humans. In the marmoset (*Callithrix jacchus*), they identified single neurons and distributed population-based codes that represented the identity of another marmoset upon viewing their face or hearing their vocalization. By revealing similarities and differences across species, the findings prompt questions about the universality of hippocampal computational principles.

Primates, including humans, have specialized cortical networks that are dedicated to processing facial features. However, individuals with prosopagnosia, a neurological condition that impairs face recognition, can identify familiar individuals on the basis of their voice or clothing (*3*). This suggests that the brain integrates various sensory elements across the anatomically localized systems that primarily process unimodal features. How does the brain accomplish such a meta-representation? Previous work established that the hippocampus fosters binding of multimodal information in humans (*4*).

A hallmark of information binding are concept cells, which respond to different pictures of the same person (for example, front or profile views of the person in different clothing and with a different facial expression) and to their spoken name, thereby encoding the concept of their identity rather than the individual perceptual features (*1*). However, although the hippocampus is an evolutionarily conserved structure, only unimodal representation of others (as in, separate populations of cells responding to either faces or voices of other individuals) had been described in nonhuman animals (*5*). This has hindered investigation of whether common neuronal mechanisms underlie the recognition of identity across species, which is important given the use of animal models in clinical research.

Tyree *et al.* found cells in the marmoset hippocampus that respond to the face and the vocalizations of familiar marmoset individuals, much as human concept cells respond to the face and the name of other humans. The presence of these cells in the marmoset, which diverged from the evolutionary path of humans nearly 40 million years ago, suggests that there is an evolutionarily rooted ability of the hippocampus to create abstract neural representations of another's identity. Ecologically relevant stimuli may be necessary for experiments to reveal such neuronal properties, which might explain previous failed attempts to detect multimodal identity representation in other nonhuman species. Indeed, marmosets, which live in close family groups, exhibit a rich acoustic communication that they use to maintain contact when foraging in forests (*6*).

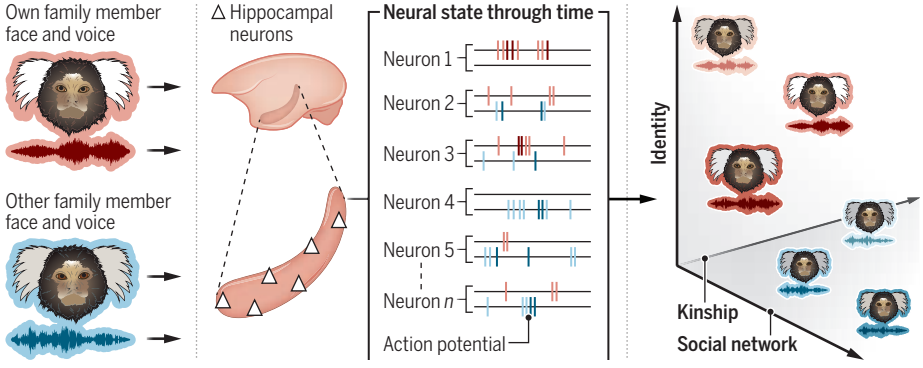
The observation of concept-like cells in a species lacking symbolic language builds on previous examples of abstraction in the macaque hippocampus (*7*, *8*). These examples include the observation that neurons encode the logical organization of recurring elements across different spaces (*7*) and form an abstract map of the relative value between different rewards (*8*). This reinforces the view that the hippocampus builds higher-level representations of elements that make up the animal's environment. Whether the concept-like cells in the marmoset could reach a semantic level of acoustic communication is an intriguing question because some primate species have adapted existing calls to signal the presence of categorically related (air or ground) but previously unseen types of predators (*9*).

In addition, Tyree *et al.* found that hippocampal cells in the marmoset encode sets of associated identities, which accounts for social networks, such as family groups. Similarly, concept cells in the human respond to conceptually related items (for example, characters from the same movie), thereby representing the learned relationship among elements (*10*). This capacity of concept cells implies that elements of an experience can be represented by combinatorics (akin to the assemblage of letters into words) by adjusting neuronal selectivity to nest individuals within larger groups.

A similar fine-grained representation of social identity was recently observed in the hippocampus of bats (*11*), whose evolutionary lineage diverged from that of humans about 95 million years ago. Wireless recordings during bat flights identified hippocam-

Encoding multimodal identity

Following an acoustic (vocalization) or a visual (face) stimulation, hippocampal neurons in the marmoset encode an abstract representation of identity. In addition, variations in the moment to moment state of the neural population encode identity, kinship and position within the social network.



Institut des Sciences Cognitives Marc Jeannerod, Centre National de la Recherche Scientifique, Bron, France. Email: sylvia.wirth@isc.cnrs.fr

pal cells that encoded the presence of specific individuals at landing sites, depending on their proximity in the social network (socially close bats were represented more often compared with socially distant bats). Activity of hippocampal cells in stationary bats was also slightly modulated by the echolocation calls of other bats, suggesting some degree of multimodal representation. In both bats and marmosets, the socially modulated neurons were anatomically dispersed in the hippocampus, confirming previous findings in humans (*1*). Thus, in phylogenetically distant species, fine-grained social signals are represented at all stages of information processing within the hippocampal circuitry, which likely facilitates the incorporation of social information in new memories.

In addition to selective concept-like cells, Tyree *et al.* identified distributed population codes for the representation of multimodal identity, which challenges the current understanding of hippocampal computational principles. These distributed codes, which have not yet been observed in the human hippocampus, enable the extraction of identity from differences in the moment-to-moment state of the neuronal population in response to multiple stimuli over time, rather than from the selectivity of single neurons for a specific stimulus (see the figure). This observation in the marmoset is supported by data from bats and rats (*12*, *13*), whose hippocampal place cells—once thought to represent a single position within a space—have been found to represent multiple positions when animals move within large spaces. Such findings suggest that the representation of place emerges from distinct states of the broader population of place cells. Overall, these observations have implications for understanding computational efficiency across different contexts and species. These findings also reveal what could be important species-specific idiosyncrasies, suggesting that in nonhuman animals, representation of information in the hippocampus may depend less on the selectivity of individual neurons than it does in humans (*14*). Whether this reflects differences in cognitive ability is not yet clear. ■

REFERENCES AND NOTES

1. R. Q. Quiroga *et al.*, *Nature* **435**, 1102 (2005).
2. T. J. Tyree *et al.*, *Science* **382**, 417 (2023).
3. B. Rossion, *Neuropsychologia* **173**, 108278 (2022).
4. B. Mišić *et al.*, *PLOS Comput. Biol.* **10**, e1003982 (2014).
5. J. Sliwa *et al.*, *Cereb. Cortex* **26**, 950 (2016).
6. C. T. Miller *et al.*, *Am. J. Primatol.* **72**, 974 (2010).
7. P. Baraduc *et al.*, *Science* **363**, 635 (2019).
8. E. B. Knudsen, J. D. Wallis, *Cell* **184**, 4640 (2021).
9. F. Wegdell *et al.*, *Nat. Ecol. Evol.* **3**, 1039 (2019).
10. E. De Falco *et al.*, *Nat. Commun.* **7**, 13408 (2016).
11. A. Forli, M. M. Yartsev, *Nature* **621**, 796 (2023).
12. T. Eliav *et al.*, *Science* **372**, eabg4020 (2021).
13. H. Zhang *et al.*, *Nat. Neurosci.* **26**, 131 (2023).
14. R. Q. Quiroga, *Hippocampus* **33**, 616 (2023).

10.1126/science.adk8413

GRAPHIC: N. CARY/*SCIENCE*

CHEMISTRY

Interconverting mirror-image molecules

A light-driven multitasking catalyst enhances chirality in molecular mixtures

By **Sojung F. Kim** *and* **Richmond Sarpong**

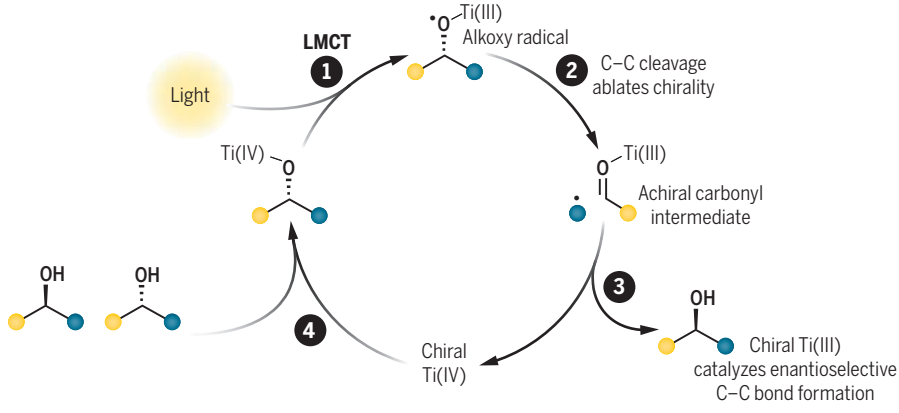
Enantiomers, though simply mirror-image molecules, can cause considerably different biological effects. Therefore, pharmacological and bioactive compounds may require separation (resolution) of one enantiomer from a racemic or scalemic mixture (equal or unequal mixtures of enantiomers, respectively). Most resolutions achieve enantiomeric enrichment by sacrificing 50% of, or chemically modifying, the material (*1*). Direct deracemization (the 100% conversion of a mixture to a single enantiomer) would be conceptually ideal. However, there are substantial barriers to realizing such a strategy, and highly selective deracemizations remain rare. On page 458 of this issue, Wen *et al.* (*2*) present a light-driven deracemization of alcohols using a single chiral titanium (Ti) catalyst that breaks and remakes carbon–carbon (C–C) bonds to achieve high levels of enantiomeric excess across a broad range of substrates by combining two enantioselective processes into one reaction. Such deracemizations could enable precise control of chirality in complex molecules.

Deracemizations face two main challenges. They increase order in the system (decrease entropy) without releasing energy (no change in enthalpy) because enantiomers are equal in energy. The result is that deracemizations are endergonic processes—energy is required to drive the reaction in the desired direction (*1*). Moreover, the principles of microscopic reversibility and detailed balance dictate that at equilibrium, forward and reverse processes occur at the same rate and by microscopically identical chemical steps in opposite directions (*3*). Any equilibrium process that converts one enantiomer to the other just as easily converts the desired enantiomer back to the undesired one, resulting in a racemic mixture. Thus, deracemizations require distinct mechanisms for the removal and formation of stereocenters (where a chiral carbon is bonded to four distinct groups in the molecule).

These thermodynamic and kinetic requirements can be met by designing redox-driven deracemizations, for example, to deracemize mixtures of indolines or tetrahydroisoquinolines (*4*, *5*), which are specific nitrogen-containing molecules with pharmacological importance. However, redox deracemizations require the consumption of a larger amount of reagents than is desirable. Photochemistry provides a nonequilibrium alternative be-

A multifunctional catalyst converts alcohols into their mirror image

A chiral titanium (Ti) catalyst selectively enriches a mixture of enantiomers. A light-mediated ligand-to-metal charge transfer (LMCT) process excites the Ti(IV)-alcohol complex and generates Ti(III) and an alkoxy radical (**1**). The C–C bond in the alkoxy radical is broken, which results in an achiral carbonyl intermediate (**2**). The generated Ti(III) complex enantioselectively catalyzes C–C bond formation to yield the desired alcohol enantiomer (**3**). The Ti(IV) catalyst is regenerated to restart the process and preferentially binds the undesired enantiomer (**4**).



cause light absorption offers both energy and access to distinct excited electronic states. The forward and reverse mechanisms can then occur across different electronic states and reaction pathways (*1*, *6*). Both excited-state energy and electron transfer have recently been leveraged to achieve photocatalytic deracemizations (*6*–*8*). However, to date, achieving multiplicative enhancement of enantioselectivity (through multiple enantioselective mechanistic steps) in a deracemization has required three separate catalysts to compartmentalize individual mechanistic steps and two distinct chiral catalysts to mediate sequential enantioselective reactions (*6*), which increases complexity and the number of required reagents.

Wen *et al.* used a single chiral Ti(IV) catalyst that enantioselectively breaks and remakes a key C–C bond that contains stereocenters through distinct mechanisms to achieve multiplicative enantioenrichment. This streamlined method does not irreversibly consume any chemical reagents, and only one Ti(IV) catalyst compound is needed. The authors leveraged ligand-to-metal charge transfer (LMCT), a relatively underutilized mode of activation by light that excites an electron from the ligand or metal-ligand bonds to an antibonding orbital of the metal (*9*). For alkoxides (deprotonated alcohols), this process results in alkoxy radicals (*10*, *11*). Crucially, high-valent metal ions can act as both LMCT catalysts and Lewis acids, meaning, they can catalyze distinct mechanistic steps, which provides the catalyst with a multitasking activity.

The chiral Ti catalyst used by Wen *et al.* preferentially binds to one enantiomer of the racemic alcohol mixture. Upon visible-light irradiation, LMCT generates a Ti(III) species and an alkoxy radical, which undergoes a C–C bond cleavage to form an achiral intermediate containing a carbonyl and a carbon radical (see the figure). This first step, which erases stereochemical information, occurs faster for the undesired enantiomer because the chiral Ti(IV) preferentially binds to it. Next, the generated Ti(III) catalyzes enantioselective C–C bond reformation from the achiral intermediate, kinetically favoring the formation of the desired enantiomer. In this way, light and LMCT mediate a Ti(IV)-Ti(III) cycle in which a single catalyst induces two mechanistically distinct enantioselective events.

Wen *et al.* obtained high levels of enantiomeric enrichment with both cyclic and acyclic alcohols with adjacent stereocenters, alcohols on fully substituted carbons, acyclic amino-alcohols, and strained four- and seven-membered rings. Moderate stereoenhancement in both the C–C bond cleavage and formation steps multiplies to yield excellent levels of enantiomeric enrichment, and the moderate enantiodifferentiation ability of the catalyst in each step is amplified in the overall transformation. The result is a highly selective deracemization of alcohols that controls the stereochemistry of C–C bonds, in an expansion from previous deracemizations that modified peripheral carbon–hydrogen (C–H) stereocenters. The ability to control the stereochemistry of C–C bonds is crucial to organic synthesis because they make up the skeleton of most organic compounds. Moreover, C–C bonds have recently been leveraged as unconventional synthetic handles for precise editing of molecular cyclic cores (*12*, *13*), enabling rapid diversification of complex molecules.

The demonstration of this streamlined photochemical deracemization on a broad range of substrates serves as the foundation for potential expansion toward deracemizations or stereoisomerizations (changing one or more stereocenters in a molecule with several stereocenters) of more complex substrates. For example, several biologically or commercially relevant chiral alcohols [such as sugars (*14*), RNA, and proteins] could serve as prime targets for selective deracemizations or stereoisomerizations. The work of Wen *et al.* makes precisely editing any desired stereocenter in complex molecules a more reachable goal. Such control could simplify the synthesis of stereochemically complex molecules by rendering lengthy, challenging resynthesis or resolution unnecessary. Deracemizations of molecules that contain other functional groups, such as amines, thiols, phosphines, or even halides, may now be possible. ■

REFERENCES AND NOTES

1. M. Huang, T. Pan, X. Jiang, S. Luo, *J. Am. Chem. Soc.* **145**, 10917 (2023).
2. L. Wen *et al.*, *Science* **382**, 458 (2023).
3. D. G. Blackmond, *Angew. Chem. Int. Ed.* **48**, 2648 (2009).
4. A. D. Lackner, A. V. Samant, F. D. Toste, *J. Am. Chem. Soc.* **135**, 14090 (2013).
5. Y. Ji, *et al.*, *J. Am. Chem. Soc.* **137**, 10496 (2015).
6. N. Y. Shin, *et al.*, *Science* **366**, 364 (2019).
7. X. Li *et al.*, *Angew. Chem. Int. Ed.* **59**, 21640 (2020).
8. C. Onneken *et al.*, *Nature* **621**, 753 (2023).
9. F. Juliá, *ChemCatChem* **14**, e202200916 (2022).
10. A. Hu *et al.*, *J. Am. Chem. Soc.* **140**, 13580 (2018).
11. L. Chang, *et al.*, *Chem. Rev.* **122**, 2429 (2022).
12. X.-Y. Yu, J.-R. Chen, W.-J. Xiao, *Chem. Rev.* **121**, 506 (2021).
13. J. Jurczyk *et al.*, *Nat. Synth.* **1**, 352 (2022).
14. Y. Wang, H. M. Carder, A. E. Wendlandt, *Nature* **578**, 403 (2020).

ACKNOWLEDGMENTS

S.F.K. acknowledges financial support from the National Science Foundation Graduate Research Fellowship Program (DGE 1752814).

Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. Email: sfkim@berkeley.edu; rsarpong@berkeley.edu

10.1126/science.adk7116

PHOTONICS

A quantum process in a laser microchip

Optical comb lasers could increase data transmission rates in telecommunications

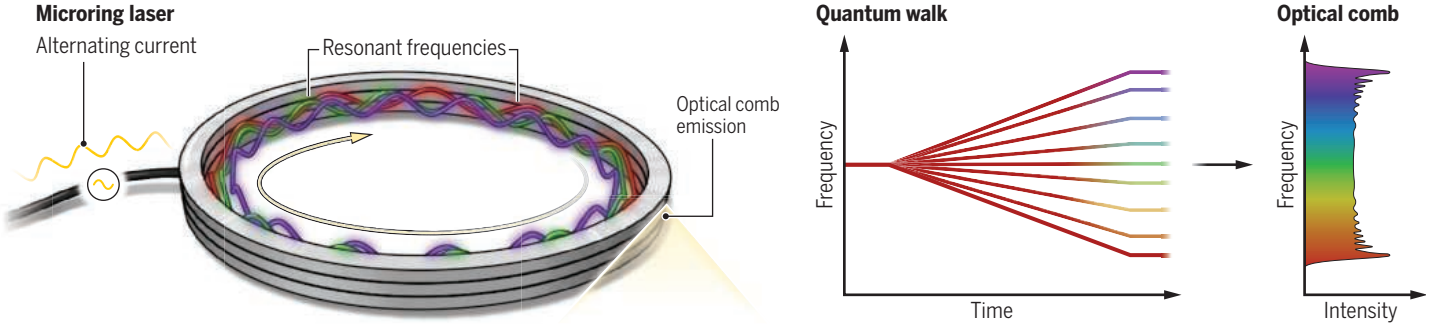
By **Giacomo Corrielli**

Optical combs are laser devices that emit hundreds of color tones simultaneously with evenly spaced frequencies. Their introduction in the late 1990s revolutionized the use of light to precisely measure time, distance, and energy (*1*). Research has focused on the miniaturization of optical combs, squeezing sophisticated tabletop setups into coin-sized microchips. This can broaden the applications of optical combs, from enabling fast optical communications to the development of new-generation telescopes (*2*). However, this size transition requires redesigning complex optical mechanisms in small devices. On page 434 of this issue, Heckelmann *et al.* (*3*) describe a process that, inspired by quantum mechanics, transforms a single-frequency (color) microchip laser into a miniaturized optical comb, which emits light with a broad spectrum composed of many frequencies. Quantum mechanics had not been previously used to describe the mechanism of an optical comb formation; in this way, the work improves the understanding of miniaturized lasers' operations.

Heckelmann *et al.* started from a known type of laser, called a quantum cascade laser, which is integrated into a microchip and emits light in the mid- to far-infrared portion of the electromagnetic spectrum (*4*). This laser is made of a tiny ring of a semiconductor alloy containing arsenide, gallium, indium, and aluminum. The microring confines and guides the light in a similar way to how metallic cables conduct electricity. When this microring is connected to a steady, direct electrical current, the electrons in the semiconductor are stimulated to quickly jump across its layers, producing a cascade of photons (i.e., light particles). As the photons circulate in the microring, they are multiplied,

A miniaturized optical comb laser based on a quantum walk

A microring laser is excited with an alternating electrical current (left). The alternating current induces coupling between the microring resonant frequencies, which results in a "quantum walk" of their frequencies that lasts a few hundred nanoseconds (middle). The process produces a plethora of colors with equally spaced frequencies, which constitutes an optical comb (right).



resulting in a single-frequency laser emission. Heckelmann *et al.* discovered that if this system is excited with an additional alternating current that oscillates at a ring resonance frequency—i.e., a frequency at which an electromagnetic wave oscillates an integer number of times around the microring—the laser emission evolves spontaneously from a single color to multiple colors in a few hundred nanoseconds (see the figure). The alternating current induces a coupling between different ring resonances and creates a "virtual" lattice of concatenated frequencies that can exchange optical energy between them. After an avalanche process, all frequencies associated with the virtual lattice are created from the initial lasing frequency, and this leads to the formation of the optical comb–like spectrum.

What makes the process described by Heckelmann *et al.* fascinating is that before stabilizing to its final shape, the light spectrum mimics the motion of a so-called quantum walk in the virtual lattice of the possible microring resonances. A quantum walk is the motion of a "walker" entity exploring multiple positions in a random fashion and following the rules of quantum mechanics (*5*). Because of the wave-like nature of quantum objects (such as electrons or photons), quantum walks are dominated by interference and diffraction effects. This leads to substantially different outcomes compared with those of their classical counterparts—i.e., the random walks—which better describe classical stochastic processes, such as the motion of living cells or the price variations of stock markets. Instead, the close analogy between quantum mechanics and classical wave phenomena allows quantum walks to serve as effective mathematical models for describing and predicting the behavior of various electromagnetic systems, including optical ones. This is the case of the experiment of Heckelmann *et al.*, where the quantum-like walker is the laser light frequency.

The application of the mathematical formalism of quantum mechanics to optics has led to important technological advances in the field of photonics. The development of guided optics (or waveguides) stands out as an example (*6*). At the same time, constructing optical devices as model systems to study new quantum phenomena has been successful and remains the foundation for the development of new-generation analog computers (photonic quantum simulators) that promise to solve specific quantum computing tasks, such as efficient drug design or modeling complex biological systems (*7*). The results of Heckelmann *et al.* demonstrate that the cross-fertilization between different disciplines—i.e., optics and quantum mechanics—can be pivotal for producing new discoveries.

"Quantum mechanics had not been previously used to describe the mechanism of an optical comb formation..."

The quantum walk comb exhibits features that make it attractive from a technological point of view. The laser emits light over a large and very stable bandwidth, with up to 100 colors in its spectrum. This is crucial, for example, for using this microchip for telecommunication purposes because each color can serve as an independent communication channel. Thus, increasing the number of colors means transmitting more information with a single device (*8*). The operation of this optical comb is fully controlled by electrical, rather than optical, signals on the microring, which substantially simplifies the operation of the device and reduces the power supply requirements for integrating the components on a single platform.

There has been tremendous growth in the development of integrated optical combs in the past 15 years, leveraging the advances of modern nanotechnology infrastructures. Different approaches, materials, and techniques are under study, and the quantum walk optical comb of Heckelmann *et al.* adds to this portfolio. Other promising solutions include the spectral broadening of ultrafast lasers (*9*) and the creation of a frequency comb through the interaction of light with nonlinear optical materials (*10*). Once these devices reach sufficient maturity in terms of comb purity and stability, degree of integration, and cost, a plethora of opportunities will open up in high-precision sensing, metrology, and photonic computing as well as autonomous navigation, telecommunication, and astronomy (*11*). ■

Istituto di Fotonica e Nanotecnologie-Consiglio Nazionale delle Ricerche (IFN-CNR), Milan, Italy.
Email: giacomo.corrielli@cnr.it

GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. S. A. Diddams, K. Vahala, T. Udem, *Science* **369**, eaay3676 (2020).
2. A. L. Gaeta, M. Lipson, T. J. Kippenberg, *Nat. Photonics* **13**, 158 (2019).
3. I. Heckelmann *et al.*, *Science* **382**, 434 (2023).
4. M. S. Vitiello, G. Scalari, B. Williams, P. De Natale, *Opt. Express* **23**, 5167 (2015).
5. W. Dür, R. Raussendorf, V. M. Kendon, H.-J. Briegel, *Phys. Rev. A* **66**, 052319 (2002).
6. S. Longhi, *Laser Photonics Rev.* **3**, 243 (2009).
7. A. Aspuru-Guzik, P. Walther, *Nat. Phys.* **8**, 285 (2012).
8. Y. Okawachi, B. Y. Kim, M. Lipson, A. L. Gaeta, *Optica* **10**, 977 (2023).
9. S. Cuyvers *et al.*, *Laser Photonics Rev.* **15**, 2000485 (2021).
10. M. H. P. Pfeiffer *et al.*, *Optica* **4**, 684 (2017).
11. L. Chang, S. Liu, J. E. Bowers, *Nat. Photonics* **16**, 95 (2022).

10.1126/science.adl0611

POLICY FORUM

TECHNOLOGY AND LAW

Artificial intelligence and interspecific law

Law could recognize nonhuman AI-led corporate entities

By **Daniel J. Gervais**[1] *and* **John J. Nay**[2]

Several experts have warned about artificial intelligence (AI) exceeding human capabilities, a "singularity" at which it might evolve beyond human control. Whether this will ever happen is a matter of conjecture. A legal singularity is afoot, however: For the first time, nonhuman entities that are not directed by humans may enter the legal system as a new "species" of legal subjects. This possibility of an "interspecific" legal system provides an opportunity to consider how AI might be built and governed. We argue that the legal system may be more ready for AI agents than many believe. Rather than attempt to ban development of powerful AI, wrapping of AI in legal form could reduce undesired AI behavior by defining targets for legal action and by providing a research agenda to improve AI governance, by embedding law into AI agents, and by training AI compliance agents.

"Law" is a human invention. Humans make the laws and the enforcement institutions. The law is composed of directives according to which, "human beings are required to do or abstain from certain actions, whether they wish to or not" [(*1*), pp. 78–79]. Humans have occasionally invited into their legal order nonhumans such as animals, rivers, and ethereal entities created by humans called "legal persons" (*2*). Yet, neither animals nor rivers have a mind or intentions that make them responsible for their behavior in the eyes of the law. Up to this point, the legal system has been univocal; it only allows humans to speak to design and use it. In the legal system, nonhuman legal subjects have necessarily instantiated their rights through human proxies; those subjects "are simply a vehicle for addressing human interests and obligations" [(*2*), pp. 591–592].

The humanness of the legal system goes further. The system allows humans (vis-à-vis other humans) to exert control over other species. Humans can "own" other species, generating rights and responsibilities, including the right to decide which members of other species live or die. Thus, the law certainly affects other species, but it is meant to regulate relations among humans about other species.

Human laws are written using language, another distinctively human invention. Only humans speak our language, which we see both as evidence of the existence of our (superior) mind and, hence, as Chomsky put it, as a "true species property" (*3*). Functionally, the abilities that have made us the only species able to design and interact with the legal system is language. Language, and its emergent concepts, allow humans to understand norms and institutions that together constitute our legal system.

Human language is no longer distinctive to humans. We have developed AI that can do as well as or better than the best humans on at least 150 cognitive tasks. A capacity to understand human language and legal norms unlocks the ability of AI to interact with the legal system. Work in recent years on contract drafting, pretrial discovery, and now more complex tasks such as understanding fiduciary obligations all point toward the conclusion that it is only a matter of time before an "understanding" of human law by AI is achieved (*4*). Recent evidence includes the ability of large language models (LLMs) to accomplish hundreds of legal reasoning tasks (*5*).

Yet, even before machines outperform lawyers at much of what lawyers do, AI may enter the legal system as legal subject. This is a catalyst for a serious conversation. Our research is not motivated by any normative claim about rights for AI or "robots," whether based on the ontological properties of advanced AI or on the direct application of a social-relational model. It is about finding a way to apply what has essentially been human law to autonomous AI capable of performing many cognitive tasks that until recently only humans could—specifically when, having taken corporate form, an AI interacts with humans, businesses, and the legal system.

CORPORATE BUT NOT HUMAN

A major law reference book refers to corporations as "artificial persons" (*6*). This is particularly apt in this context because that is precisely what is happening; AI could take corporate form. The laws of several jurisdictions in the United States and elsewhere do not always explicitly require humans at the helm of a legal person. Some laws do require the existence of a board of directors, which, by interpretation, some could say must be composed of humans, but not all jurisdictions require such a board. The laws are even laxer in the case of Limited Liability Companies (LLCs). Overall, nothing generally prevents an AI from managing the affairs of a corporate entity. By law, corporate entities need not have human owners or managers.

The idea of a "zero-member LLC" is not new [e.g., (*7*)]. What has changed is that it is now possible from a technological standpoint. And the Uniform Limited Liability Company Act (ULLCA) explicitly provides for the possibility. Though it notes that absence of any members leads to dissolution of an LLC, the ULLCA did not make that provision mandatory (*8*). Even if only one US state allowed zero-member LLCs, that entity could operate nationwide under the so-called "internal affairs doctrine," according to which courts look to the law of the state of incorporation for rules governing the internal affairs of a corporate entity.

For instance, under state laws that made it the first US state to authorize the formation of LLCs, Wyoming could become the home of the first zero-member LLC. Although its corporate laws allowing the formation of zero-member LLCs predate AI, the state government is aware of the possibility and has not taken steps to prevent it. Indeed, the state would like to attract AI investment, like many other jurisdictions. Envision a zero-member LLC operating autonomously: taking orders for specific products online or via email; contacting suppliers; arranging for shipping and payment; managing feedback, returns, and complaints, all without any direct human involvement in the process.

If such an AI-operated LLC did exist, what options would the law have? The law

[1]Law School, Vanderbilt University, Nashville, TN, USA. [2]CodeX–The Center for Legal Informatics, Stanford University, Stanford, CA, USA. Email: daniel.gervais@vanderbilt.edu; john.j.nay@gmail.com

could not easily attach legal consequences to the autonomous acts of the AI—acknowledging that "autonomy" here may well be limited (but so it is generally also for humans). The law could perhaps attach responsibility to humans who launched the LLC, but this might require a new legal toolkit, as humans are generally shielded from liability for acts of a corporate entity that they create or control. Moreover, it is not inconceivable that an AI itself could apply to establish a new LLC, especially if it is already operating as one. Courts would still have all the tools they now use when a corporate entity is held liable: damages, seizure of assets, and dissolution come to mind. Courts could issue orders such as injunctions, but the decision to comply would be made by the AI.

There is another option: making zero-member LLCs illegal in all jurisdictions, but that would require a massive legislative effort worldwide and arguably runs counter to the ethos of attracting and growing technology industries that animate many jurisdictions around the world. Thus the emergence of zero-member LLCs should catalyze the legal system to adapt to autonomous AI agents. That said, the creation of AI-operated zero-member LLCs is difficult to reconcile with theories of the firm, which are anthropocentric and instrumental in nature. Those frameworks are based on the benefits that the firm provides to humans, just as, say, intellectual property is designed to ensure human progress (*9*).

CLEAR LEGAL SUBJECT

AI now has the capabilities to trade in currency. Although trading in fiat currencies (such as the US dollar, euro, yen, etc.) is highly regulated, AI can trade in digital currencies that settle on blockchains. The machine could thus pay humans to perform tasks that require interfacing with the physical world—at least until they can be performed efficiently by robots. Digital currencies can be decentralized by relying on distributed blockchain technology. That can then allow for AI to operate in a decentralized way, leading to the formation of decentralized autonomous organizations (DAOs) that make it (even) harder to regulate.

Having an AI operate as an LLC, or other legal entity, provides a clear legal subject. This has two benefits: It makes AI a target for legal action for compensatory damages paid by the corporate entity; and it provides a clearer research agenda for machine learning researchers to improve AI governance.

Using the legal system to create zero-member LLCs is a way to test its limits. But many corporations are already increasingly managed by AI. Many human resource functions, such as filtering of applications, have been delegated to AI. Copywriters are being replaced by LLMs, whose role in advertising is already prominent and poised to grow exponentially. Many aspects of financial planning can be performed by LLMs leveraging tools.

As AI moves up the cognitive tree and can perform ever more complex functions—in many cases surpassing humans—why would it not enter the C-suite? Soon, many parts of a chief technology officer's job (and possibly a chief financial officer's) will likely be done by AI. Lawyers already rely on many AI-powered tools. It won't be long before a corporation can ask AI for legal advice, which may, with time, be better than many human lawyers can offer.

"Machines are first introduced as…'copilots.' …But the purpose of a copilot is to be able to take over at some point."

This is happening fast because tens of billions of dollars are being spent to build more bridges, at ever increasing levels, between human and machine cognition. Machines are first introduced as assistants, or "copilots" (*10*). This "assist first, replace later approach" is natural because it enables a machine to be gradually trusted, put in place, and then AI capabilities improve through further research advances and through additional training data collected from real-world deployments. But the purpose of a copilot is to be able to take over at some point. At what stage does the machine no longer need feedback from humans specifically, rather than automated feedback or feedback from other AI systems? This is what AI is trending toward.

REGULATION AND RESEARCH

Because AI can be used in almost any field of activity—e.g., investing, legal services, power grid management, warfare, health care—the regulation of AI as a single, new regulatory target is awkward. Each governmental agency should instead consider the impact of AI and, in our opinion, the real possibility that autonomous AI agents will exist among humans.

Lawmakers face an unprecedented challenge: regulating AI that can perform cognitive tasks that until recently only humans could. What happens, as a matter of law, when another "species" interacts with us and makes "real-world" decisions—not through human proxies, but directly—and does all this "intelligently"? There is a real risk that AI technology will slip beyond human control. The regulatory equation to solve is that, even if AI machines can behave like humans—and in some cases much faster—we cannot regulate them exactly as humans (*11*). However, machine learning and computer science research more broadly may be able to unlock law-following tendencies in AI that, when paired with interspecific law, could enable AI governance.

Within the current paradigm, there appears to be a path toward embedding law-following into LLM-powered AI agents (*12*). For example, humans can train an LLM to accurately predict which real-world actions are more consistent with a given legal standard, e.g., fiduciary duties. The resulting LLM can be used to monitor, and/or influence, a primary AI agent, and it could be used as a reward model (in reinforcement learning parlance). If the latter, humans could train a "student" LLM with reinforcement learning against that reward model "teacher." The student may then exhibit behavior more consistent with the legal standard it was trained against. This would represent a step toward LLM-powered AI agents aligned with human law. To be clear, there are many situations involving human agents that ultimately require human adjudication of what is legal in that context; this does not change with AI agents. Embedding a deeper understanding of law into AI agents seeks to address the vast majority of day-to-day actions and situations but will never handle the highly ambiguous, or the edge cases that require a human court opinion.

More broadly, data generated by legal processes and the tools of law (methods of lawmaking, statutory interpretation, contract drafting, applications of standards, and legal reasoning) can be leveraged by researchers to train AI systems to learn methods for internalizing what one could label the "spirit" of the law, that is, the robust specification of often inherently vague human goals. Asimov's laws were a parable for why brittle rules (if-then statements) do not suffice. Our proposal is to instead focus AI research on legal standards and the spirit of directives.

We can continue to see the legal system with humans sitting alone at the top, allowing only physical things that we designate (specific lakes and rivers) or ethereal things created out of thin air (corporations) to exist as potential right holders. Beyond the normative case that one could make against this, the very idea with the advanced forms of AI that will emerge in the next few years is that we may want them to be part of our legal system so that we can track their actions, apportion blame, put guardrails around behavior, and guide AI research toward building lawful artificial "consciences." Making

legal compliance a core component of the AI training process and deploying an adversarial automated monitoring requirement is a potential path forward.

The two main enforcement tools of the legal system are financial penalties and imprisonment. As Stuart Russell noted, "[i]f we were to imprison the robot for nonpayment, why would it care?" [(*13*), p. 126]. The same could be said of the imposition of financial penalties, though an AI might "care" if it operates an LLC and has to obtain general liability insurance. But imagine an AI in an offshore jurisdiction where it may be easier to operate, trading only in digital currency. What grip would human law have on that entity? Instead of running from that, but having it happen, we can run toward bringing AI into the fold in a way that we can track everything it does and subject it to every law that a human must comply with, plus potentially more law specific to AI. But the key is that everything a human is subject to, the AI is as well.

The reality is that most people obey the law most of the time without any enforcement. Humans internalize behavioral norms consistent with legal standards. We may understand that if we were to ignore all laws—knowing that there is only so much the state can do to enforce them—a general state of chaos might emerge. For a variety of hard-to-express reasons, humans generally view reasonable compliance as desirable.

By contrast, machines are amoral. They can predict nuances of human morality, but their way of "internalizing" morals is different. Machine thinking can be much more powerful than human thinking, but it is devoid of key human characteristics such as interoception, the sense of what is happening inside our bodies. The challenge is to get the machines to "want" to comply without relying on an inherent inner compass that many humans seem to be endowed with through some combination of nature and nurture. The challenge goes beyond aligning AI with a set of ethical rules—however those are determined—which risks creating a second set of rules for AI that is not democratically determined through law-making processes.

The challenge is to align AI machines with the law on the books. We use the zero-member LLC as an exemplar of why this is desirable because an LLC should comply with the law (*14*). This is potentially more likely to happen, in our view, if we can make the case that we are willing to bring advanced AI into the legal system after it has had sufficient legal compliance training and is subject to sufficient adversarial monitoring and live AI-driven (but human interpretable) feedback. This AI-driven adversarial monitoring is part of how we keep the primary AI systems in line with ever-evolving law.

We can imagine a scenario in which the government requires any sufficiently advanced AI deployment to be accompanied by a certified AI compliance agent and, separately, a form of government-run AI auditor or police leveraging similar AI legal compliance technology for a final line of public defense. Our focus is on catalyzing AI research that can help prevent illegal behavior of AI agents through private and/or public initiatives. With the speed of digital intelligence, a focus on proactive automated crime prevention is critical. But in the inevitable situations where AI systems don't live up to the ideal behavior because of wildly out-of-distribution situations in which they find themselves, the relevant corporate entity should be liable for paying damages to those that prove harm to a court. To enable this, frameworks could be extended that require minimum business liability insurance policies to be held. This is common practice for many areas of business. And, of course, where a human can be found liable for misuse of AI, they should be.

"The challenge is to get the machines to 'want' to comply without relying on an inherent inner compass..."

BRING AI INTO THE LEGAL FOLD

Interspecific law will happen, but it is impossible to predict where on the spectrum we will end up. At one end, interspecific law means slightly adapting corporate law to the operation of corporate entities with humans only partly in control. This is imminent because AI is already in control of many corporate activities. At the other end, it means adapting the legal system to everyday interactions with autonomous, intelligent entities in cases where prosaic tools of the legal system, like post hoc enforcement, are unlikely to work in the traditional way.

Some will dismiss the idea of letting machines enter the legal system as subjects as nonsense. To those readers, three things. First, they could do so soon, at least indirectly, using zero-person LLCs where there is no human governing a legal entity, as explained above. Second, if we don't proactively wrap AI agents in legal entities that must obey human law, then we lose considerable benefits of tracking what they do, shaping how they do it, and preventing harm. Third, if the point is that machines will never have powerful agency well beyond human oversight: We hope you are right!

There seems to be another choice. We could try to put a hard stop on AI. Some scientists have warned that developing AI more advanced than humans is extremely risky. In our view, this hard stop will likely not happen. Capitalism is en marche. There is too much innovation and money at stake, and societal stability historically has relied on continued growth.

AI replacing most human cognitive tasks is a process that is already underway and seems poised to accelerate, although the possibility of a major public opinion backlash that would limit the scope and depth of replacement of humans cannot be excluded (*15*). This means that our options are effectively limited: Try to regulate AI by treating the machines as legally inferior, or architect AI systems to be law-following, and bring them into the legal fold now with compliance tendencies baked into them and their AI-powered automated legal guardrails. Research directions described here suggest that this may be possible. We think it may be desirable. ■

REFERENCES AND NOTES

1. H. L. A. Hart, *The Concept of Law* (Clarendon, 1994).
2. R. L. Cupp Jr., *Fla. Law Rev.* **465**, 506 (2017).
3. J. Chen, *Wash. Univ. Law Q.* **73**, 1263 (1995).
4. J. Nay (2023); https://law.stanford.edu/publications/large-language-models-as-fiduciaries-a-case-study-toward-robustly-communicating-with-artificial-intelligence-through-legal-standards/.
5. N. Guha *et al.*, LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models (2023); https://arxiv.org/abs/2308.11462.
6. *Black's Law Dictionary* defines an "artificial person" as a "nonhuman entity that is created by law and is legally different owning its own rights and duties. AKA juristic person and legal person."
7. S. J. Bayern, *Northwest. Univ. Law Rev.* **108**, 1485 (2014).
8. https://www.uniformlaws.org/committees/community-home?CommunityKey=bbea059c-6853-4f45-b69b-7ca2e49cf740. As of 2023, 20 US states have neither enacted the ULLCA nor introduced legislation to do so. It is worth nothing that the 2013 amendments to the ULLCA (RULLCA) do seem to require at least one person as member, for example Section 20, and Section 201.
9. B. R. Cheffins, "The Rise and Fall (?) of the Berle-Means Corporation" (2018); https://doi.org/10.17863/CAM.27945.
10. See, for example, https://github.com/features/copilot.
11. T. Arnold, M. Scheutz, *Ethics Inf. Technol.* **20**, 59 (2018).
12. J. J. Nay, *Northwest. J. Tech. Intell. Prop.* **20**, 309 (2023).
13. S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking, 2019).
14. C. O'Keefe, Law-Following AI 1: Sequence Introduction and Structure, AI Alignment Forum (2022); https://www.alignmentforum.org/posts/NrtbF3JHFnqBCztXC/law-following-ai-1-sequence-introduction-and-structure.
15. J. Stilgoe, *Science* 10.1126/science.adk0176 (2023).

ACKNOWLEDGMENTS

J.J.N. is the founder and CEO of Nomos AI and a cofounder and the chairman of Brooklyn Investment Group.

10.1126/science.adi8678

Wissahickon Creek served as a source of specime and inspiration for the Morris sisters' work.

BOOKS *et al.*

SCIENCE LIVES

The lost scientists

A tale of two 19th-century researchers reveals how professionalization led to women's exclusion from science

By **Sarah Boon**

In *Mischievous Creatures: The Forgotten Sisters Who Transformed Early American Science*, historian Catherine McNeur tells the story of siblings Margaretta and Elizabeth Morris, two 19th-century American scientists whose contributions to their respective fields have long been underappreciated. The pair were independently wealthy and remained single throughout their lives, which allowed them to focus their time on science. Despite their connections in the scientific community and the privileges afforded by the vast resources at their disposal, the sisters struggled to be taken seriously, particularly as science transformed from a personal vocation into a professional pursuit.

Mischievous Creatures: The Forgotten Sisters Who Transformed Early American Science
Catherine McNeur
Basic Books, 2023. 432 pp.

The 1800s, writes McNeur, were "a golden age for popular science" in the United States. Young men and women were rigorously trained in scientific topics at school, and the public regularly attended scientific lectures. Being a successful 19th-century scientist relied heavily on letter-writing, publication, and networking. Although this opened doors for Margaretta and Elizabeth, women were increasingly excluded from scientific endeavors as science became professionalized.

Margaretta and Elizabeth received scientific training through a combination of school lessons and private tutoring. The latter enabled them to build relationships with a number of well-regarded scientists, including the English botanist Thomas Nuttall, the American entomologist Thomas Say, and the French naturalist Charles Alexandre Lesueur. Each sister focused on a different discipline: Margaretta pursued entomology, while Elizabeth chose botany.

Margaretta was ambitious and shared her findings widely, interacting regularly with her cousin, the chemist Robert Hare, and other well-known scientists, including Charles Darwin. She published papers in leading journals and was one of the first women admitted to the American Association for the Advancement of Science (AAAS) and to the Academy of Natural Sciences (ANS). She was best known for her work on cicadas, finding that Brood X nymphs lived on plant roots underground, sucking sap for sustenance in a manner that damaged trees.

Elizabeth, in contrast, was happy to work behind the scenes, although she too interacted regularly with a wide range of scientists, including the American botanist Asa Gray and the Swiss American biologist Louis Agassiz. She sent botanical samples to collectors who requested them, connected up-and-coming botanists with established ones in her extensive network, and wrote prolifically for the same journals in which Margaretta published, although she chose to write under a pseudonym. "Anonymity was a protective shield," notes McNeur, "especially for women at a time when writing under their own names might be perceived as improper or even vulgar."

Despite their training and contributions, the sisters were often slighted by their male peers. Margaretta's ANS membership, for example, was listed as "exempt"—a designation that excluded her from meetings and prohibited her from voting. Women, observes McNeur, were frequently considered to be merely dabbling in fields in which men were making real contributions, and their accomplishments were gradually erased as scientific gatekeeping expanded and men began to dominate the ranks of "professional" scientists.

This erasure, in turn, has long hampered the creation of community for women scientists, obscuring the achievements of their predecessors and making it difficult for historians of science to document women's contributions. The letters and other writings of early women scientists, notes McNeur, were often either discarded or discovered only in the correspondence of their male peers.

Although it never quite delivers on the subtitle's promise to reveal how Margaretta and Elizabeth Morris "transformed early American science"—the women undoubtedly made considerable contributions to their respective fields, but their efforts were not necessarily transformative to science itself—*Mischievous Creatures* nonetheless offers a fascinating portrait of 19th-century science in the United States from a little-considered perspective. The sisters' story illuminates the historical role of women in science and the challenges they faced, many of which may unfortunately still resonate with women scientists today. ■

The reviewer is a freelance writer and editor from Vancouver Island, Canada. Email: snowhydro1@gmail.com

10.1126/science.adk4874

COMMUNICATION

Writing without words

A computational linguist probes the relationship between graphical symbol systems and language

By **Andrew Robinson**

Having trained as a linguist at the Massachusetts Institute of Technology, Richard Sproat moved into computational linguistics at AT&T Labs in the 1980s, then into academic linguistics, and is today a researcher on text normalization and speech recognition at Google. He is therefore intimately familiar with symbols of all kinds, as demonstrated in his latest ambitious work, *Symbols*. The book is the first systematic study of graphical symbol systems, ranging from the imagery found in Paleolithic cave paintings, through ancient and contemporary writing systems employing both phonetic and logographic symbols, to modern language-independent symbols such as meteorological icons and emoji.

Sproat begins with an inquiry sent to him nearly 20 years ago by premodern historian Steve Farmer, who was then working with Indologist Michael Witzel on a project concerning the ancient Indus script, dating from ~2600 BCE. First discovered in the Indus Valley in the late 19th century, the script remains undeciphered. Farmer's question to Sproat was "deceptively simple," as Sproat recalls: Is it "possible to distinguish statistically any linguistic character string represented by a fairly large corpus of texts from nonlinguistic symbol chains?" His answer was "no."

In 2004, having analyzed the limited Indus corpus and its apparent symbolic patterns, Farmer, Sproat, and Witzel published a paper arguing that the symbols were not a writing system but rather some sort of nonlinguistic system without phonetic symbols (*1*). This predictably upset many Indus scholars, six of whom, led by Rajesh Rao, replied with a paper in *Science* in 2009 (*2*). A debate ensued, lasting for several years. "As is often the case, the people who disagree with you the most can often be the ones who most force you to understand the issues better," admits Sproat in his preface.

Symbols is, in a sense, the product of this controversy. The role of language in writing systems preoccupies it. Must writing have a phonetic element, Sproat wonders. He identifies two chief schools of thought about how to define writing—inclusivism and exclusivism—as specified by sinologist John DeFrancis (*3*). According to DeFrancis, "inclusivists" classify as writing all systems of graphic symbols that can convey "some amount of thought" without using phoneticism. Examples include mathematical symbols, musical notation, and emoji. Meanwhile, "exclusivists" reject such systems and limit writing to symbols that can convey "any and all thought." Mesopotamian cuneiform, European alphabets, and Chinese characters, all of which are dependent on phoneticism, meet exclusivists' criteria.

This debate naturally leads Sproat, an exclusivist, to consider how symbols came to express language or, in other words, how writing evolved in ancient times. Unfortunately, hard evidence for writing's origins is lacking. The earliest known writing certainly dates from Mesopotamia, ~3300 BCE, but whether this influenced through diffusion the invention of later Egyptian hieroglyphs, Chinese characters, and Mesoamerican scripts remains unclear. Moreover, although the earliest Mesopotamian symbols certainly involved accountancy, there is no clear evidence of how these nonlinguistic symbols evolved into the later cuneiform symbols capable of expressing language, despite theories offered by many scholars.

What is clear is that civilization preceded writing. In Sproat's shrewd words, "Writing is the product of civilization, but it is not a *necessary* product of civilization." We know this from archaeological excavation of Mesopotamian cities—Uruk, for example—that surely required some bureaucracy in order to flourish, which they did before they had access to writing.

He hypothesizes that writing that expressed language arose from accountancy symbols that did not, as follows: "Writing evolved in an *institutional* context in which symbols were effectively *dictated*, so that the user of the symbol system gradually came to associate the symbols with sounds."

As for the futuristic vision of a "universal" writing system in which symbols convey "any and all thought" independent of the world's spoken languages—a proposal first advocated by philosopher and mathematician Gottfried Leibniz in the 17th century—this will remain a dream, according to Sproat. "Language is still our most versatile form of communication," he writes, "and unless Elon Musk's prediction that his Neuralink implants will make language and speech obsolete comes true, it seems destined to remain that way." ■

The reviewer is the author of *The Story of Writing: Alphabets, Hieroglyphs and Pictograms* (Thames & Hudson, 2007) and *The Indus: Lost Civilizations* (Reaktion Books, 2021).
Email: andrew@andrew-robinson.org

Symbols: An Evolutionary History from the Stone Age to the Future
Richard Sproat
Springer Cham, 2023. 235 pp.

REFERENCES AND NOTES

1. S. Farmer *et al.*, *Electron. J. Vedic Stud.* **11**, 19 (2004).
2. R. P. N. Rao *et al.*, *Science* **324**, 1165 (2009).
3. J. DeFrancis, *Visible Speech: The Diverse Oneness of Writing Systems* (Univ. of Hawaii Press, 1989).

10.1126/science.adk9395

SCIENCE, SEX, AND GENDER

PHOTO: LEV DOLGACHOV/ALAMY STOCK PHOTO

PODCAST

Lab Hopping: A Journey to Find India's Women in Science
Aashima Dogra and Nandita Jayaraj
India Viking, 2023. 302 pp.

The discrimination and harassment faced by women scientists in the West have been increasingly discussed and debated in recent years; however, the challenges faced by women researchers in non-Western contexts remain underexplored. This week on the *Science* podcast, Aashima Dogra and Nandita Jayaraj examine how gender intersects with other identities and with social expectations in India's scientific community. bit.ly/46bln7j

10.1126/science.adl1585

Flooding in Brazil's Rio Grande do Sul highlights the need for policies that anticipate the risks of climate change.

PHOTO: TIAGO COELHO/BLOOMBERG VIA GETTY IMAGES

Edited by **Jennifer Sills**

Brazil undermines flood resilience

In September, heavy rains caused substantial flooding in Brazil's state of Rio Grande do Sul (*1*). In the past 2 years, similar events have occurred in the states of São Paulo (*2*), Rio de Janeiro (*3*), Minas Gerais, Espírito Santo, Tocantins, Bahia, Piauí, Pará, and Maranhão (*4*, *5*). Heavy rains and flooding events are expected to increase as global warming continues (*6*, *7*). Therefore, President Luiz Inácio Lula da Silva's administration must update the country's laws to account for the effects of climate change.

In 2021, during the previous administration, Brazil's President Jair Bolsonaro signed Bill 2510/2019 into Law 14,285/2021 (*8*). The legislation removed the federal restrictions (enumerated in Law 12,651/2012) that were in place to protect from development strips of land ranging from 30 to 500 m in width along watercourses in urban areas (*9*). Before the changes, for example, a 30-m strip was protected on each side of a stream under 10 m in width (*9*). The new guidelines permit local governments to allow houses or other infrastructure in areas near watercourses (*8*). As a result, many municipalities have already declared that buildings can be built closer to the banks of rivers and streams; some have established distances as small as 5 m (*10*). Riverbanks are the areas most vulnerable to extreme weather events [e.g., (*3*)]. Many existing houses and other structures are already very close to watercourses, posing risks to humans in the case of floods.

Instead of increasing the risks caused by extreme events, Brazil should be working to increase the land's resilience to flooding. Riparian vegetation—which will also be extirpated under the new law—protects watercourses (*11*), regulates flow, and helps limit damage from extreme climatic events. In addition to passing a new law to revoke Law 14,285/2021 (*12*), Brazil urgently needs policies to restore occupied and paved areas along streams and rivers (*12*). Brazil's laws should strengthen environmental protection and reaffirm the country's commitments to international environmental treaties.

Valter M. Azevedo-Santos[1,2,3]*, Vanessa S. Daga[4], Lívia H. Tonella[5], Renata Ruaro[6], Marlene S. Arcifa[7], Philip M. Fearnside[8], Tommaso Giarrizzo[2,9]
[1]Programa de Pós-Graduação em Biodiversidade, Ecologia, e Conservação, Universidade Federal do Tocantins, CEP 77500-000, Porto Nacional, Tocantins, Brazil. [2]Grupo de Ecologia Aquática, Espaço Inovação do Parque de Ciência e Tecnologia Guamá, Belém, Pará, Brazil. [3]Faculdade Eduvale de Avaré, Avaré, São Paulo, Brazil. [4]Deltona, FL, USA. [5]Faculdade Serra do Carmo, Palmas, Tocantins, Brazil. [6]Programa de Pós-graduação em Ciência e Tecnologia Ambiental, Universidade Tecnológica Federal do Paraná, Curitiba, Paraná, Brazil. [7]Departamento de Biologia, Universidade de São Paulo, 14040-901 Ribeirão Preto, São Paulo, Brazil. [8]Instituto Nacional de Pesquisas da Amazônia, CEP 69067-375, Manaus, Amazonas, Brazil. [9]Instituto de Ciências do Mar (LABOMAR), Universidade Federal do Ceará, 60165-081 Fortaleza, Ceará, Brazil.
*Corresponding author.
Email: valter.ecologia@gmail.com

REFERENCES AND NOTES

1. "Chuvas no RS: corpo é encontrado na orla de rio de Porto Alegre," *Correio Braziliense* (2023) [in Portuguese].
2. P. Chatterjee, "São Paulo: Dozens killed as deadly storms hit Brazilian coast," *BBC News* (2023).
3. "Rio Paraíba do Sul transborda e alaga ruas de Paraíba do Sul," *G1* (2023) [in Portuguese].
4. P. Jackson, "Rain storm claims 100 lives in north-east Brazil," *BBC News* (2022).
5. G. Alecrim, D. Porto, "Chuvas deixam mais de 113 mil pessoas fora de casa em oito estados," *CNN Brasil* (2022) [in Portuguese].
6. N. S. Debortoli, P. I. M. Camarinha, J. A. Marengo, R. R. Rodrigues, *Natural Haz.* **86**, 557 (2017).
7. M. Zachariah *et al.*, "Interplay of climate change-exacerbated rainfall, exposure and vulnerability led to widespread impacts in the Mediterranean region," *World Weather Attribution* (2023).
8. Brazil, Presidência da República, "Lei Nº 14.285, de 29 de dezembro de 2021," Diário Oficial da União, 246, Section 1, p. 5 (2021); https://in.gov.br/en/web/dou/-/lei-n-14.285-de-29-de-dezembro-de-2021-370917982 [in Portuguese].
9. Brazil, Presidência da República, "Lei Nº 12.651, de 25 de Maio de 2012" (2012); https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12651.htm [in Portuguese].
10. Encruzilhada do Sul, "Lei Complementar Nº 18, 15 de Dezembro de 2022" (2022); https://www.encruzilhadadosul.rs.gov.br/portal/leis_decretos/2258/ [in Portuguese].
11. R. Lowrance *et al.*, *BioScience* **34**, 374 (1984).
12. V. M. Azevedo-Santos *et al.*, *Land Use Pol.* **131**, 106721 (2023).

10.1126/science.adl0891

Hope for funding biodiversity efforts

The climate crisis and the biodiversity crisis are intertwined (*1*, *2*). Climate change can accelerate biodiversity loss, and the associated ecosystem degradation undermines ecosystem resilience and reduces climate change mitigation by reducing carbon sequestration (*2*). However, because decision-makers and the public devote more attention to climate change (*2*), there is a large disparity between resources allocated to climate change mitigation and those allocated to biodiversity conservation and restoration (*3*, *4*). The European Union (EU) spent about €201 billion (USD211 billion) between 2014 and 2020 on climate change mitigation (*4*) and plans

to invest another €118 billion (USD124 billion) between 2021 and 2027 (*5*), an average of almost €23 billion (USD24 billion) per year. This sum far exceeds the amount spent on biodiversity by the entire international community, estimated to be between USD4 billion and USD10 billion per year (*4*, *6*). By funding biodiversity efforts as well as climate change solutions, international coalitions can deliver a more effective response to both challenges.

The international community has recently taken steps toward better funding biodiversity efforts. At the Global Environment Facility's Seventh Assembly in Vancouver, Canada, in 2023, 186 countries agreed to create the Global Biodiversity Framework Fund to increase investment in restoration and renewal of nature (*7*), and two countries—Canada and the United Kingdom—committed to initial contributions of CAD200 million (USD161 million) and £10 million (USD12 million), respectively (*7*). The fund promises to mobilize resources from public, private, and philanthropic sources, with a focus on biodiversity and ecosystem sustainability (*7*).

This biodiversity fund will allow investment in crucial protection of natural ecosystems, which will also serve as a cost-effective way to mitigate climate change. For example, Brazil's "conservation units" (protected areas) have all been created to preserve biodiversity, yet they provide important climate benefits by maintaining carbon stocks and recycling water (*8*). Although the immediate goal is to protect existing carbon stocks in intact ecosystems, global land vegetation can sequester an extra 13.7 gigatons of carbon per year if additional management practices are adopted (*9*). Wild animals and their ecosystem roles are key components of natural climate solutions that can potentially sequester 6.5 gigatons of carbon per year (*10*).

Biodiversity (*4*) and ecosystem function (*2*) must be included in responses to climate change, including the discussions about resource allocation during 2023 United Nations Climate Change Conference, or Conference of the Parties (COP) Climate 28, which begins on November 30, as well as at COP Biodiversity 16 and COP Climate 29, both scheduled for 2024. Focusing on these issues equally is imperative to ensure a balanced distribution of economic resources, reflecting their comparable importance and impact on humanity.

Cássio Cardoso Pereira[1]*, Geraldo Wilson Fernandes[1,2], Daniel Negreiros[2], Walisson Kenedy-Siqueira[2], Stephannie Fernandes[3], Philip M. Fearnside[4]
[1]Programa de Pós-Graduação em Ecologia, Conservação e Manejo da Vida Silvestre, Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. [2]Knowledge Center for Biodiversity, Belo Horizonte, MG, 31270-901, Brazil. [3]Department of Global and Sociocultural Studies, Florida International University, Miami, FL, USA. [4]Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, 69067-375, Brazil.
*Corresponding author.
Email: cassiocardosopereira@gmail.com

REFERENCES AND NOTES

1. H.-O. Pörtner *et al.*, *Science* **380**, eabl4881 (2023).
2. C. C. Pereira *et al.*, *Nat. Conserv. Bulgaria* **53**, 157 (2023).
3. J. Reed *et al.*, *One Earth* **3**, 753 (2020).
4. N. Pettorelli *et al.*, *J. Appl. Ecol.* **58**, 2384 (2021).
5. European Commission, "The EU dedicates €118 billion to climate change projects through cohesion policy" (2023); https://ec.europa.eu/regional_policy/whats-new/panorama/2023/10/10-11-2023-the-eu-dedicates-eur118-billion-to-climate-change-projects-through-cohesion-policy_en.
6. E. B. Barbier *et al.*, *Science* **360**, 486 (2018).
7. "New global biodiversity fund launched in Vancouver," Global Environment Facility (2023); https://www.thegef.org/newsroom/press-releases/new-global-biodiversity-fund-launched-vancouver.
8. P. M. Fearnside, *Biodivers. Conserv.* **30**, 1199 (2021).
9. Z. Sha *et al.*, *Commun. Earth Environ.* **3**, 8 (2022).
10. O. J. Schmitz *et al.*, *Nat. Clim. Chang.* **13**, 324 (2023).

10.1126/science.adk7806

The Asiatic cheetah's road to extinction

The expansion of road networks in Iran has substantially affected the critically endangered Asiatic cheetah (*1*–*4*). Since 2004, at least 23 Asiatic cheetahs have been killed in vehicle collisions (*4*). The worldwide Asiatic cheetah population is now estimated to be fewer than 20 adults, all of whom live in Iran (*5*). The country must take immediate action to protect the species from extinction.

Ten cheetahs, one of which was pregnant, have been killed on Semnan-Mashhad Road alone (*6*), which is located at the periphery of the Touran Biosphere Reserve, supposedly the last remaining sanctuary for Asiatic cheetahs (*2*). In 2018, Iran established a 3-km fenced area near the reserve to prevent cheetahs from accessing the road (*6*). After the fence was built, cheetah road accidents ceased until 2022 (*6*). However, because the fence only covered part of the reserve's border and no monitoring was implemented, it is unclear whether the temporary halt in road-kills resulted from the intervention. Iran has proposed several additional collision mitigation measures, including extending the fence from 3 km to 36 km on both sides of the road, retrofitting culverts under the road as wildlife corridors, installing roadway lighting, and deploying speed control cameras (*7*).

Insufficient funding, exacerbated by the devaluation of the Iranian rial (*8*, *9*), poses a substantial obstacle to minimizing the number of Asiatic cheetahs killed on roads. Nongovernmental organizations have launched donation campaigns but struggle to raise the USD2 million that the fence extension proposal estimates will be required (*10*). The Ministry of Roads and Urban Development of Iran has pledged cooperation with the Iran Department of Environment (DOE) to fund about 15 km of fencing (*7*). The Conservation of Asiatic Cheetah Project was established to reverse the decline of the Asiatic cheetah (*11*), but suspended operations after difficulties collaborating with Iranian government officials. Reinstating it could improve conservation efforts. Moving forward, Iran should track the success of its interventions to identify the most effective conservation strategies. However, the most immediate need is stronger commitments from DOE and nongovernmental organizations to implement and evaluate the effectiveness of every possible intervention that could save this species from global extinction.

Alireza Mohammadi[1]*, Ali Ranjbaran[2], Mohammad S. Farhadinia[3,4], José Vicente López-Bao[5], Anthony Paul Clevenger[6]
[1]Department of Environmental Science and Engineering, Faculty of Natural Resources, University of Jiroft, Jiroft, Iran. [2]Tehran, Iran. [3]Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Kent, UK. [4]Department of Biology, University of Oxford, Oxford, UK. [5]Biodiversity Research Institute, Spanish National Research Council, Oviedo University, Mieres, Spain. [6]Western Transportation Institute, Bozeman, MT, USA.
*Corresponding author.
Email: armohammadi1989@gmail.com

REFERENCES AND NOTES

1. J. Parchizadeh *et al.*, *Curr. Biol.* **28**, R1141 (2018).
2. E. M. Moqanaki, S. A. Cushman, *Anim. Conserv.* **20**, 29 (2017).
3. M. Ahmadi *et al.*, *Divers. Distrib.* **23**, 592 (2017).
4. M. S. Farhadinia *et al.*, *Biodivers. Conserv.* **26**, 1027 (2017).
5. H. Akbari, "There are fewer than 20 cheetahs left in the wild of the country: Warning about the high statistics of wildlife trafficking in Iran," *Iranian Labor News Agency* (2014); https://www.ilna.ir/fa/tiny/news-1357732 [in Farsi].
6. H. Akbari, "At least 17 Asiatic cheetahs live in Turan National Park: We still hope for the pregnancy of 'Iran' again," *Hakim Mehr News Agency* (2023); http://hakimemehr.ir/fa/news/76114 [in Farsi].
7. H. Akbari, "The restoration of the historical path of movement of fencing is examined more closely," *Islamic Republic News Agency* (2023); https://www.irna.ir/news/85102859/[in Farsi].
8. L. Khalatbari *et al.*, *J. Sci.* **362**, 1255 (2018).
9. J. Parchizadeh, J. L. Belant, *Nature* **610**, 448 (2022).
10. Nazr Nature Association, "People's campaign of Aman Yoz crossing" (2023); https://www.nazretabiat.ir/campaign/122/securing-cheetah-road [in Farsi].
11. L. Hunter *et al.*, *Cat News* **46**, e11 (2007).

COMPETING INTERESTS

J.V.L.-B. is a member of the Canid Specialist Group of the International Union for Conservation of Nature Species Survival Commission and a member of the Scientific Committee advising the Spanish Ministry for the Ecological Transition and the Demographic Challenge, both unpaid advisory roles.

10.1126/science.adk7900

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

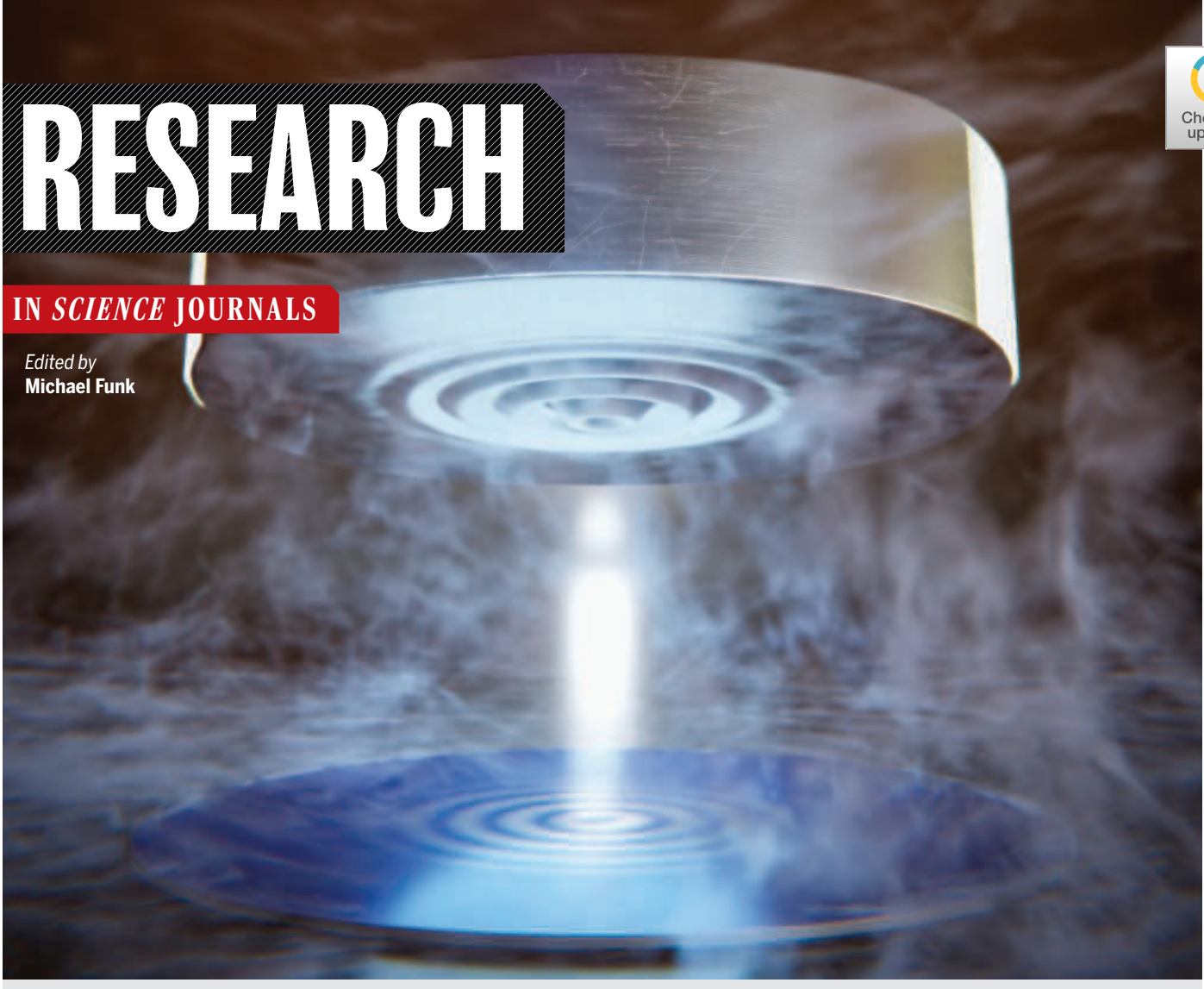


ILLUSTRATION: M. BORST & SCIXELES

SPINTRONICS

Controlling spin-waves with a superconductor

The ability to control spin and charge transport on a chip has revolutionized information technology. For spintronics, the collective spin excitations of magnetic materials, known as spin-waves, have emerged as a promising platform that can offer new functionalities because of their wave nature. However, control of the spin-waves has remained a formidable challenge. Borst *et al.* show that the diamagnetism of a superconductor can be used to shape the magnetic environment that governs the transport of spin-waves in a magnetic thin film. Magnetic imaging showed how the diamagnetism leads to spin-waves with strongly altered, temperature-tunable wavelengths. Control of spin-wave transport using the diamagnetic response of a superconducting gate will be important for developing device applications. —ISO Science, adj7576, this issue p. 430

Artist's depiction of spin-wave transport within a superconductor

PLANT SCIENCE

Initiating the Casparian strip

Plant roots contain a lignin-based diffusion barrier called the Casparian strip that controls nutrient balance. Gao *et al*. found that a family of dirigent proteins controls the synthesis of lignin at the Casparian strip. The authors established a two-stage process of synthesis in which the dirigent proteins initiate lignin biosynthesis in plasma membrane patches. After this, gaps between the patches are filled in by the previously described Schengen pathway. The work extends our understanding of complex polysaccharide patterning in the plant extracellular space. —MRS

Science, adi5032, this issue p. 464

IMMUNOLOGY

Translation-dependent IL-17 generation

Interleukin-17 (IL-17)–producing CD4$^+$ helper T (T_H17) cells mediate chronic inflammation in autoimmune diseases. Revu *et al.* uncovered a nuclear role for the phospholipid PIP_2 in IL-17 production by T_H17 cells. During the differentiation of human naïve CD4$^+$ T cells into T_H17 cells in vitro, PIP_2 was enriched in nuclei in a manner dependent on the kinase PIP5K1α. Nuclear PIP_2 facilitated the translation of *IL17A* mRNA and IL-17 protein production. CD4$^+$ T cells from multiple sclerosis patients treated with a PIP5K1α inhibitor secreted less IL-17A than untreated cells. —JFF

Sci. Signal. (2023)
10.1126/scisignal.abo6555

SOLID STATE PHYSICS

Electron-induced lattice softening

In solid state materials, changes in the crystal lattice are often accompanied by changes in the electronic system. Whether the lattice or the electrons is the primary driver of a transition may, however, be difficult to ascertain. Noad *et al*. measured the Young's modulus in the extremely clean material Sr_2RuO_4 as it underwent an electronic (Lifshitz) transition. The researchers found a large drop in the Young's modulus at the transition, suggesting that conduction electrons drive a nonlinear elastic response in this material. —JS

Science, adf3348, this issue p. 447

ENZYMOLOGY

Metal switch expands activity

DNA polymerases, like most enzymes, are specific for their native substrates, deoxyribonucleotides, and have carefully tuned active site chemistry that can be challenging to adapt. Lelyveld *et al*. performed time-resolved x-ray crystallography on a DNA polymerase variant that accepts a 3′-amino sugar, catalyzing the formation of phosphoramidate linkages instead of the natural phosphodiester bonds. They found that the two metal ions responsible for catalysis in the native enzyme were disrupted, and only a single divalent metal ion was involved in catalysis in this variant enzyme. The authors used this insight to propose trivalent rare earth metal ions as an alternative cofactor. The addition of scandium(III) yielded the most proficient phosphoramidate–DNA polymerase variant. —MAF

Science, adh5339, this issue p. 423

DIFFERENTIATION

Decoding human gut endocrine cell fate

Enteroendocrine cells (EECs) reside in the epithelium of the digestive tract and produce various hormones involved in metabolism. The generation of different EEC lineages is governed by a dedicated network of transcription factors. However, given the low efficiency of EEC specification from adult stem cells, it has been difficult to elucidate the components of this regulatory network. Lin *et al*. used an optimized human small intestinal organoid culture system to perform an unbiased, systematic screen for transcription factors that regulate EEC differentiation. The screen implicated ZNF800 as a key factor exerting a dominant repressive role in controlling the endocrine transcription factor network. This work showcases the use of optimized human organoids for CRISPR-based functional screens, paving the way for the identification of additional regulators in human gut physiology and pathophysiology. —SMH

Science, adi2246, this issue p. 451

INFECTIOUS DISEASES

Unraveling congenital ZIKV syndrome

Pregnant women infected with Zika virus (ZIKV) can vertically transmit the virus to their fetuses, and this has been linked to congenital ZIKV syndrome (CZS). Understanding how ZIKV infection affects prenatal and neonatal development is critical for human health. Saron *et al*. and Moadab *et al*. used cynomolgus and rhesus macaque models, respectively, to understand how maternal ZIKV infection influences fetal development. Saron *et al*. found that preexisting immunity to dengue virus, a related flavivirus, worsened the severity of CZS. Moadab *et al*. found that maternal ZIKV infection alone was sufficient to reduce infant size and affect social behaviors. Together, these studies shed light on the impacts of ZIKV infection on fetal and early life development. —CNF and CM

Sci. Transl. Med. (2023) 10.1126/scitranslmed.adh0043, 10.1126/scitranslmed.add2420

IN OTHER JOURNALS

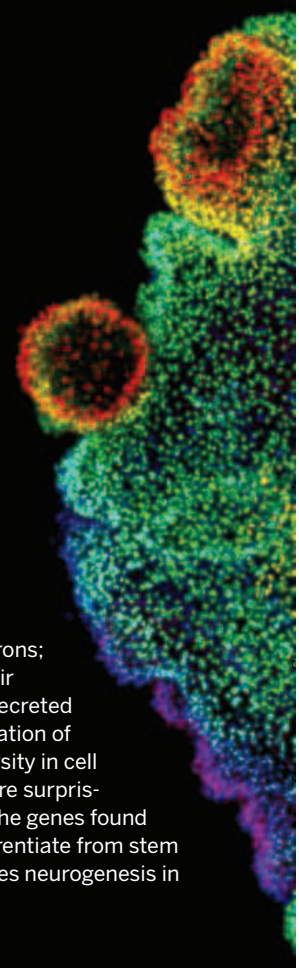
Edited by **Caroline Ash** *and* **Jesse Smith**

EVOLUTION

Prehistory of neurons

Free-living marine placozoans are considered the simplest animals on Earth. Their bodies resemble a disk and consist of six to nine cell types. They move by ciliary beating, consume food particles through engulfment, and reproduce by fission. However, Najle *et al.* have shown that these tiny, unassuming creatures may hold the secret to the evolution of the complex nervous systems of higher animals. Placozoans do not have neurons; instead, the collective behavior of their cells is controlled by neuropeptides secreted by peptidergic cells. Detailed investigation of these cells revealed unexpected diversity in cell types and functional specificities. More surprisingly, these cells expressed many of the genes found in evolved nervous systems and differentiate from stem cells through a pathway that resembles neurogenesis in higher animals. —DJ

Cell (2023) 10.1016/j.cell.2023.08.027



The disc-shaped aggregates of cells called placozoans are among the most basal of animals.

REPRODUCTIVE HEALTH

Pregnancy and job security

Pregnancy loss is a seldom discussed public health issue that affects women and their partners' health and well-being. At the population level, psychosocial stressors such as economic downturns or natural disasters can increase risks of adverse pregnancy outcomes. However, at the individual level, the stress from becoming unemployed may also increase a woman's risk of a stillbirth or miscarriage. Using longitudinal, representative data from 40,000 households in the UK, Di Nallo and Köksal examined the risks of job losses to pregnancy outcomes. They found that involuntary job losses of the pregnant woman or her partner increased the odds of pregnancy losses, whereas expected, non-involuntary job interruptions did not. —EEU

Hum. Reprod. (2023) 10.1093/humrep/dead183

MECHANICAL BONDS

A chain of infinities

Catenanes are interlocked molecular rings akin to the linked segments in a

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

MOLECULAR BIOLOGY

Toward one-step gene editing

Programmable DNA targeting has improved rapidly over the past decade, primarily as a result of the development of CRISPR-Cas systems, which allow precise cleavage of target sequences. However, this approach relies on cellular DNA repair pathways for the subsequent gene-editing step, which introduces heterogeneity that represents a safety concern. In a Perspective, Tang and Sternberg propose retroelement-based gene editing as a safer alternative to CRISPR-Cas approaches. Retroelements are endogenous DNA segments that use the host cell's RNA polymerase to generate an RNA copy of themselves, which is then turned into complementary DNA by a reverse transcriptase and incorporated into the genome elsewhere. Recent studies have provided information on the mechanisms underlying this process and raise the hope that they can be exploited for programmable DNA targeting in the future. —SAL

Science, adi3183, this issue p. 370

ANTHROPOLOGY

Another menopausal primate

Menopause occurs in all known human societies; however, it is not common to all mammals and has so far only been observed in humans and a few toothed whale species. Wood *et al*. looked at demographic and endocrine data in a long-studied population of chimpanzees in Uganda and found clear evidence for menopause in females living past the age of 50 (see the Perspective by Cant). Unlike the case for humans and toothed whales, however, postreproductive chimps in this population are not involved in the raising of related offspring, suggesting that a different process is driving its development. —SNV

Science, add5473, this issue p. 416; see also adk7119, p. 368

NEUROSCIENCE

Concept cells in nonhuman primates

There are numerous animal studies showing that single neurons encode the identity of conspecifics from social signals in different sensory modalities such as olfaction, vision, or audition. However, there has been no demonstration in a nonhuman animal that these unimodal signals are integrated into a cohesive cross-modal representation of individual identity. Tyree *et al*. performed single-neuronal recordings in marmosets presented with pictures and sounds from conspecifics (see the Perspective by Wirth). A subpopulation of neurons responded selectively to both the face and voice of individual animals, much like concept neurons in the human medial temporal lobe. Furthermore, animal identity could be successfully decoded from the neuronal population activity, along with aspects of social relationships such as being a family member. —PRS

Science, adf0460, this issue p. 417; see also adk8413, p.372

PHOTONICS

Stabilizing frequency combs

The generation of stable and controllable optical frequency combs has important implications for applications in metrology and precision spectroscopy. Heckelmann *et al*. propose and demonstrate a quantum walk laser in synthetic frequency space (see the Perspective by Corrielli). The technique allows for the development of a frequency comb laser based on a coherent walk in frequency space combined with a fast-gain medium. Also applicable to optical wavelengths, the approach provides a flexible route to generating stable and broadband frequency combs across a range of wavelengths for precision sensing and communication applications. —ISO

Science, adj3858, this issue p. 434; see also adl0611, p. 374

ORGANIC CHEMISTRY

Enriching alcohols coming and going

Asymmetric catalysis generally relies on biasing a single step in a reaction to favor one of two possible mirror-image products. This approach places a high premium on ligand optimization. Wen *et al*. report a light-driven deracemization of alcohols in which a single chiral titanium catalyst differentiates stereochemistry in both a photoinitiated bond cleavage step and a subsequent bond reformation step (see the Perspective by Kim and Sarpong). Neither step alone is exceptionally selective, but the cumulative combination leads to enantiomeric ratios ranging from 90:10 to 99:1 across a variety of substrates. —JSY

Science, adj0040, this issue p. 458; see also adk7116, p. 373

QUANTUM GASES

The dynamics of an unusual transition

Cooling two-dimensional many-body systems to low temperatures causes them to undergo the so-called Berezinskii-Kosterlitz-Thouless (BKT) transition to a superfluid phase. The BKT transition differs from the more conventional ones in three dimensions, and although it has been studied extensively for equilibrium systems, its dynamic version is relatively unexplored. Sunami *et al*. studied the dynamics of a two-dimensional Bose gas of rubidium atoms after a quench across the BKT transition. To create the quench, the researchers split the superfluid gas in half, causing each half to undergo a transition to the normal phase. Matter-wave interferometry revealed universal scaling laws in accordance with theory. —JS

Science, abq6753, this issue p. 443

TUMOR IMMUNOLOGY

A fine-tuned balance

Intratumoral exhausted $CD8^+$ stem- or progenitor-like T cell (T_{prog}) and terminally differentiated T cells (T_{term}) mediate antitumor immunity and response to immunotherapy. Sun *et al*. investigated how exhausted $CD8^+$ T cells are transcriptionally regulated in murine tumor models, finding that the transcription factors BCL6 and BLIMP1 reciprocally regulate T cell fates. BCL6 promoted the long-term persistence of T_{prog} cells in response to TGF-β–SMAD2 signaling by repressing the T_{term} transcriptional program. Conversely, interleukin-2–STAT5 signaling repressed BCL6, resulting in BLIMP1 transcription and subsequent T_{term} differentiation. These findings identify a regulatory network that controls $CD8^+$ T cell exhaustion in cancer, which could be targeted to alter the balance of T_{prog} and T_{term} cells and improve responses to immunotherapy. —HMI

Sci. Immunol. (2023) 10.1126/sciimmunol.adh1306

SOLID STATE PHYSICS

Electron-induced lattice softening

In solid state materials, changes in the crystal lattice are often accompanied by changes in the electronic system. Whether the lattice or the electrons is the primary driver of a transition may, however, be difficult to ascertain. Noad *et al*. measured the Young's modulus in the extremely clean material Sr_2RuO_4 as it underwent an electronic (Lifshitz) transition. The researchers found a large drop in the Young's modulus at the transition, suggesting that conduction electrons drive a nonlinear elastic response in this material. —JS

Science, adf3348, this issue p. 447

ENZYMOLOGY

Metal switch expands activity

DNA polymerases, like most enzymes, are specific for their native substrates, deoxyribonucleotides, and have carefully tuned active site chemistry that can be challenging to adapt. Lelyveld *et al*. performed time-resolved x-ray crystallography on a DNA polymerase variant that accepts a 3′-amino sugar, catalyzing the formation of phosphoramidate linkages instead of the natural phosphodiester bonds. They found that the two metal ions responsible for catalysis in the native enzyme were disrupted, and only a single divalent metal ion was involved in catalysis in this variant enzyme. The authors used this insight to propose trivalent rare earth metal ions as an alternative cofactor. The addition of scandium(III) yielded the most proficient phosphoramidate–DNA polymerase variant. —MAF

Science, adh5339, this issue p. 423

DIFFERENTIATION

Decoding human gut endocrine cell fate

Enteroendocrine cells (EECs) reside in the epithelium of the digestive tract and produce various hormones involved in metabolism. The generation of different EEC lineages is governed by a dedicated network of transcription factors. However, given the low efficiency of EEC specification from adult stem cells, it has been difficult to elucidate the components of this regulatory network. Lin *et al*. used an optimized human small intestinal organoid culture system to perform an unbiased, systematic screen for transcription factors that regulate EEC differentiation. The screen implicated ZNF800 as a key factor exerting a dominant repressive role in controlling the endocrine transcription factor network. This work showcases the use of optimized human organoids for CRISPR-based functional screens, paving the way for the identification of additional regulators in human gut physiology and pathophysiology. —SMH

Science, adi2246, this issue p. 451

INFECTIOUS DISEASES

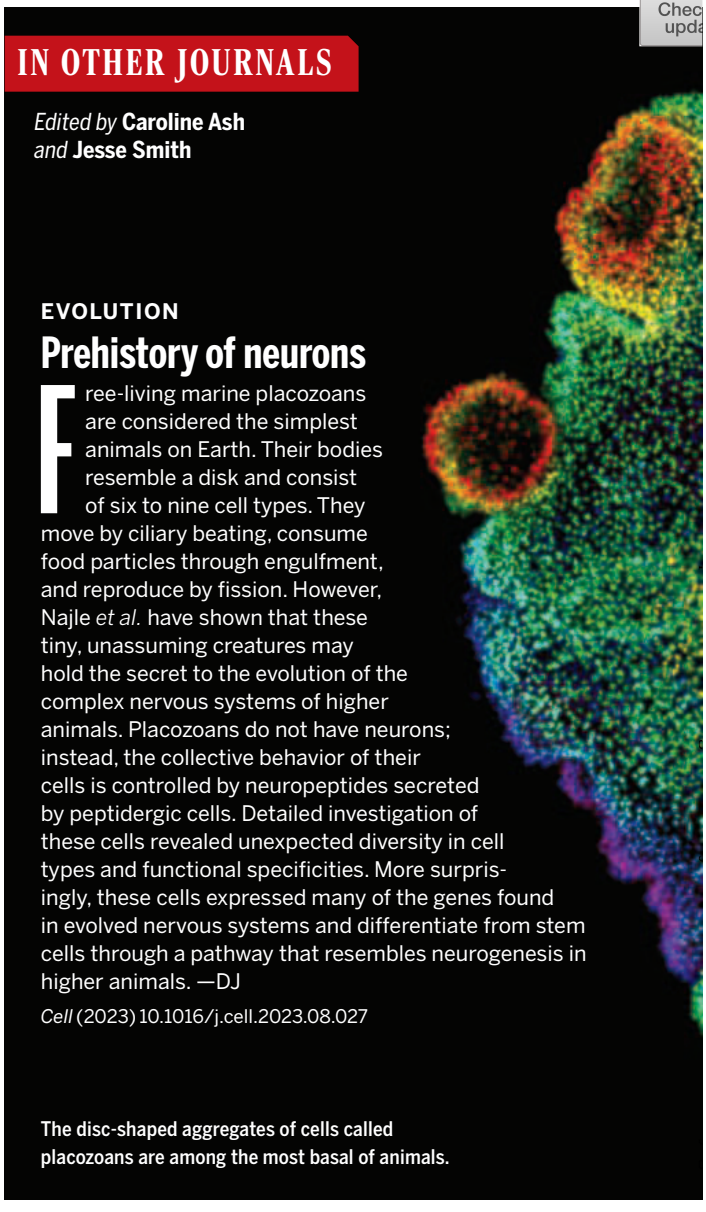
Unraveling congenital ZIKV syndrome

Pregnant women infected with Zika virus (ZIKV) can vertically transmit the virus to their fetuses, and this has been linked to congenital ZIKV syndrome (CZS). Understanding how ZIKV infection affects prenatal and neonatal development is critical for human health. Saron *et al*. and Moadab *et al*. used cynomolgus and rhesus macaque models, respectively, to understand how maternal ZIKV infection influences fetal development. Saron *et al*. found that preexisting immunity to dengue virus, a related flavivirus, worsened the severity of CZS. Moadab *et al*. found that maternal ZIKV infection alone was sufficient to reduce infant size and affect social behaviors. Together, these studies shed light on the impacts of ZIKV infection on fetal and early life development. —CNF and CM

Sci. Transl. Med. (2023) 10.1126/scitranslmed.adh0043, 10.1126/scitranslmed.add2420

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**



EVOLUTION

Prehistory of neurons

Free-living marine placozoans are considered the simplest animals on Earth. Their bodies resemble a disk and consist of six to nine cell types. They move by ciliary beating, consume food particles through engulfment, and reproduce by fission. However, Najle *et al.* have shown that these tiny, unassuming creatures may hold the secret to the evolution of the complex nervous systems of higher animals. Placozoans do not have neurons; instead, the collective behavior of their cells is controlled by neuropeptides secreted by peptidergic cells. Detailed investigation of these cells revealed unexpected diversity in cell types and functional specificities. More surprisingly, these cells expressed many of the genes found in evolved nervous systems and differentiate from stem cells through a pathway that resembles neurogenesis in higher animals. —DJ

Cell (2023) 10.1016/j.cell.2023.08.027

The disc-shaped aggregates of cells called placozoans are among the most basal of animals.

REPRODUCTIVE HEALTH

Pregnancy and job security

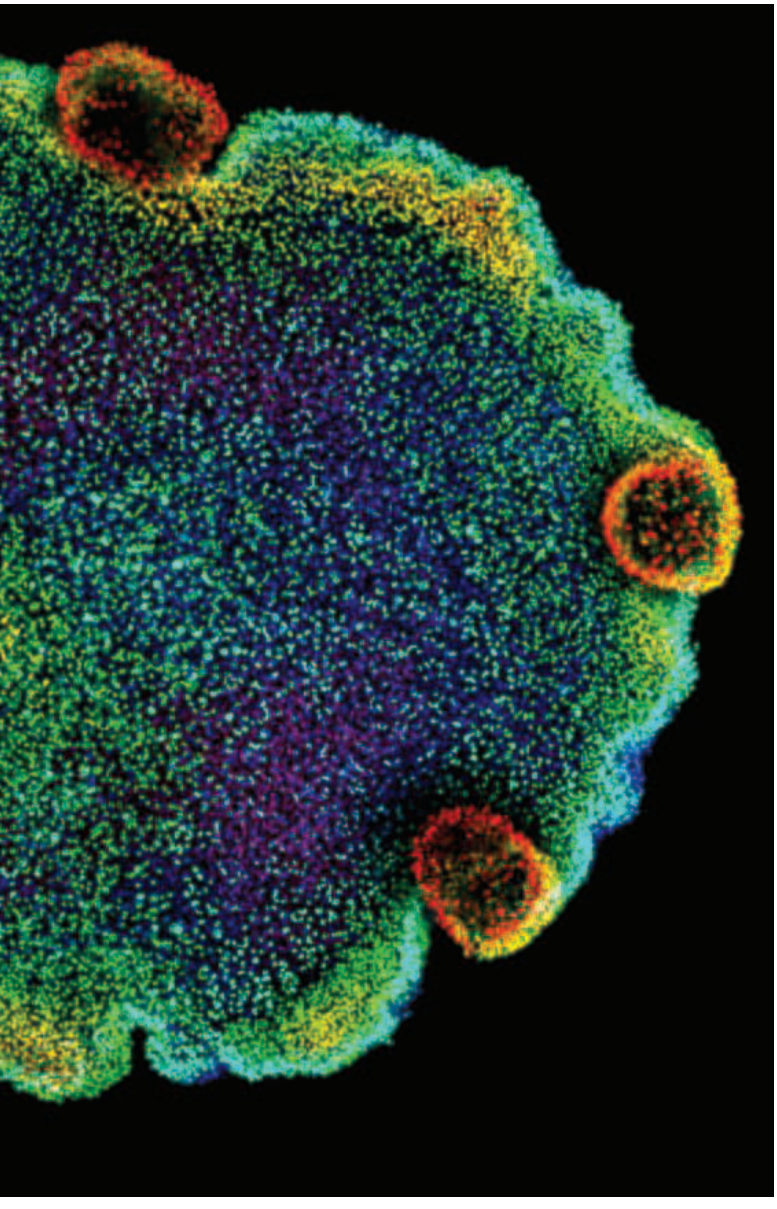
Pregnancy loss is a seldom discussed public health issue that affects women and their partners' health and well-being. At the population level, psychosocial stressors such as economic downturns or natural disasters can increase risks of adverse pregnancy outcomes. However, at the individual level, the stress from becoming unemployed may also increase a woman's risk of a stillbirth or miscarriage. Using longitudinal, representative data from 40,000 households in the UK, Di Nallo and Köksal examined the risks of job losses to pregnancy outcomes. They found that involuntary job losses of the pregnant woman or her partner increased the odds of pregnancy losses, whereas expected, non-involuntary job interruptions did not. —EEU

Hum. Reprod. (2023) 10.1093/humrep/dead183

MECHANICAL BONDS

A chain of infinities

Catenanes are interlocked molecular rings akin to the linked segments in a



macroscopic chain. Recent research has focused on methods to extend catenane chains to form polymeric materials. Liu *et al.* have now added a twist—literally—by designing ligands to curve around silver ions. They prepared chains in which the catenanes are twisted into figure-eight shapes, or perhaps more appropriately, infinity symbols (∞) extending to infinity. The actual polymers are of course finite and were characterized by crystallography, absorption spectroscopy, and electron microscopy. —JSY

Angew. Chem. Int. Ed. (2023)
10.1002/anie.202314481

GEOLOGY

Recreating early crust

The processes that created Earth's earliest crust more than 4 billion years ago continue to be debated. Hastie *et al.* turned to analog experiments to better constrain the conditions under which the early crust formed. By melting material that the authors believe represents the composition of Earth's ancient crust, they discovered that it could not form above 50 kilometers depth. To generate melts at this depth requires subduction or subduction-like processes and suggests that at least some form of plate tectonics was operating then on the early Earth. —BG

Nat. Geosci. (2003)
10.1038/s41561-023-01249-5

NEUROSCIENCE

Reward expectation in fruit flies

Foraging animals need to make rapid decisions during changing conditions. Operant matching, in which choices between options are divided in proportion to the rewards received, is an ubiquitous decision-making strategy adapted to dynamically changing environments. However, little is known about the underlying neurobiological mechanisms. To find out more, Rajagopalan *et al*. applied behavioral analysis, optogenetics, and computational modeling to the fruit fly, *Drosophila melanogaster*. A foraging task was designed in which choices made by individual fruit flies could be monitored, and it confirmed that the flies' behavior adhered to operant matching. Further, the authors found that flies' synaptic plasticity allowed expectation of reward through the protocerebral anterior medial dopaminergic neurons. —PRS

Proc. Natl. Acad. Sci. U.S.A. (2023)
10.1073/pnas.2221415120

BONE MARROW

Reservoir bones

Within bone marrow, blood vessel endothelial cells can secrete factors that are required for bone homeostasis and hematopoiesis. While surveying the taxonomy of bone marrow cells by single-cell RNA sequencing, Iga *et al*. identified a distinctive, and previously unknown, subtype of capillary endothelial cell within the femurs of mice. These were located specifically at the epiphyseal region of the long bone, a section mostly ignored when analyzing the constituents of bone marrow. In addition to producing type I collagens that strengthen bones, the epithelial cells may also form a niche for hematopoietic stem cells. —SHR

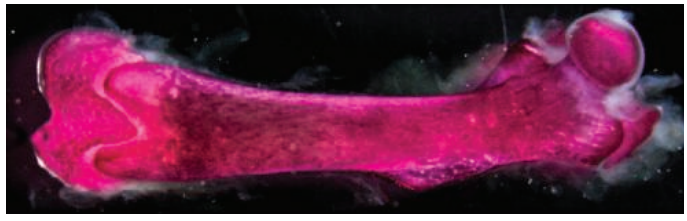
Nat. Cell. Biol. (2023)
10.1038/s41556-023-01240-7

OPTICS

Imaging by ray shifting not bending

Imaging objects with conventionally shaped glass lenses or thinner versions in the form of metalenses comprising engineered arrays of subwavelength resonators both produce an inverted image. Each method works on the premise of bending light waves, so precise alignment of imaging components is required to produce a sharp, upright image. Liu *et al.* describe an ultrathin reciprocal lens in which imaging is based on laterally shifting the light rays rather than bending them. Demonstrating the principle in a photonic crystal slab, the single reciprocal lens eliminates the need for alignment to produce an upright image. The ray-shifting mechanism provides an alternate route for the development of optical components for imaging and wavefront shaping. —ISO

Phys. Rev. X (2023)
10.1103/PhysRevX.13.031039



Alizarin red staining of infant mouse femur to observe ossification and blood vessel development

PHOTO: IGA *ET AL.*, *NATURE CELL BIOLOGY* 2023

RESEARCH ARTICLE SUMMARY

ANTHROPOLOGY

Demographic and hormonal evidence for menopause in wild chimpanzees

Brian M. Wood*, Jacob D. Negrey, Janine L. Brown, Tobias Deschner, Melissa Emery Thompson, Sholly Gunter, John C. Mitani, David P. Watts, Kevin E. Langergraber*

INTRODUCTION: It is not obvious why selection should favor menopause or the continued survival of individuals that can no longer reproduce. Among mammals, substantial numbers of post-reproductive females living under natural conditions in the wild have only been observed in humans and a few whale species. The rarity of this trait makes it both interesting and difficult to study. Data from our close primate relatives are especially valuable for the reconstruction and causal modeling of human life history evolution. In this study, we combined demographic and hormonal data to investigate post-reproductive life spans and their underlying physiological mechanisms in chimpanzees (*Pan troglodytes schweinfurthii*), who, along with bonobos, are humans' closest living relatives.

RATIONALE: We examined the mortality and fertility rates of 185 female chimpanzees in the Ngogo community of wild chimpanzees in Kibale National Park, Uganda, from 21 years of observation (1995–2016). We calculated the demographic measure PrR (post-reproductive representation), representing the fraction of adult life spent in a post-reproductive state. Human menopause, the nonpathological and permanent cessation of ovarian function resulting from the depletion of ovarian follicles, is reflected in increasing levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) and decreasing levels of ovarian steroid hormones (estrogens and progestins). To assess whether Ngogo females undergo humanlike menopause, we analyzed age-associated trends in five hormones measured in 560 urine samples from 66 females of varying reproductive status and age (range: 14 to 67 years).

RESULTS: As in other chimpanzee populations and humans, fertility declined after age 30, and no births were observed after age 50. Unlike other chimpanzee populations, but as in humans, it was not unusual for Ngogo females to live past the age of 50 (N = 16 females). The observed PrR value was 0.195, indicating that a female who reached adulthood (age 14) was post-reproductive for about one-fifth of her adult life, around half as long as human hunter-gatherers. Hormonal measures show that Ngogo females experience a reproductive transition similar to that of humans, characterized by increasing levels of FSH and LH and declining levels of estrogens and progestins as they undergo menopause.

CONCLUSION: Menopause ends reproduction around the age of 50 in both humans and wild chimpanzees. Substantial PrR has not been previously observed in any wild primate p[opu]lation, chimpanzees included. One exp[la]nation for this discrepancy is that substantial PrR could be a temporary response to unusually favorable ecological conditions at Ngogo, including low levels of predation, high food availability, and successful between-group competition. A second possibility is that substantial PrR is an evolved, species-typical trait in chimpanzees, which has not been observed elsewhere owing to recent negative human impacts, especially disease epidemics. The grandmother hypothesis suggests that older females could evolve to live past their reproductive years to help increase their daughters' fertility or their grandoffsprings' survival. This is unlikely to apply to chimpanzees, whose aged females generally live apart from their daughters, as daughters leave their natal groups at adulthood. In the context of female-biased dispersal, a more relevant theory may be the reproductive conflict hypothesis, which highlights the fact that after migrating into a new group, females become increasingly related to other group members as they age and face competition with younger females for limited breeding opportunities. The oldest females might stop reproducing in order to limit the inclusive fitness costs of that competition. The grandmother and reproductive conflict hypotheses are not mutually exclusive alternatives, and both may be required to explain why all human societies have higher PrR than documented here for chimpanzees. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: brianwood@anthro.ucla.edu (B.M.W.); kevin.langergraber@asu.edu (K.E.L.)
Cite this article as B. M. Wood *et al.*, *Science* **382**, eadd5473 (2023). DOI: 10.1126/science.add5473

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.add5473

PHOTOS: (LEFT TO RIGHT) DAVID P. WATTS, KEVIN E. LANGERGRABER AND JOHN KEVIN IFF

Three post-reproductive female chimpanzees. From left to right are MARL (died at age 69), MAR (died at 64), and Sutherland (still living at age 61).

RESEARCH ARTICLE

ANTHROPOLOGY

Demographic and hormonal evidence for menopause in wild chimpanzees

Brian M. Wood[1,2]*, Jacob D. Negrey[3], Janine L. Brown[4], Tobias Deschner[5,6], Melissa Emery Thompson[7], Sholly Gunter[8,9], John C. Mitani[10], David P. Watts[9], Kevin E. Langergraber[11]*

Among mammals, post-reproductive life spans are currently documented only in humans and a few species of toothed whales. Here we show that a post-reproductive life span exists among wild chimpanzees in the Ngogo community of Kibale National Park, Uganda. Post-reproductive representation was 0.195, indicating that a female who reached adulthood could expect to live about one-fifth of her adult life in a post-reproductive state, around half as long as human hunter-gatherers. Post-reproductive females exhibited hormonal signatures of menopause, including sharply increasing gonadotropins after age 50. We discuss whether post-reproductive life spans in wild chimpanzees occur only rarely, as a short-term response to favorable ecological conditions, or instead are an evolved species-typical trait as well as the implications of these alternatives for our understanding of the evolution of post-reproductive life spans.

In most wild vertebrates, the period of survival past the age of last reproduction is short, but the evolution of long life spans in humans is associated with a substantial post-reproductive period. In this study, we investigated whether a population of wild chimpanzees with long life expectancy experience menopause and exhibit significant post-reproductive survival. Demographic and behavioral data have been collected from the Ngogo community of wild chimpanzees in Kibale National Park, Uganda, since 1995. We examined demographic data covering the years 1995–2016, including 1611 chimpanzee risk years (i.e., the cumulative time that individuals in this study were demographically monitored) for measures of female mortality and fertility (n = 185 females). To assess post-reproductive survival at Ngogo, we calculated rates of fertility and survivorship across all ages, identified females that lived well beyond their last births, and calculated the post-reproductive representation (PrR) statistic (*1*), which represents the fraction of female adult years lived in a post-reproductive state.

In humans, reproductive cessation occurs by means of menopause, the permanent non-pathological age-associated cessation of ovarian function resulting from depletion of the lifetime supply of ovarian follicles (*2*). However, there are many other potential causes of sterility that can manifest with age, such as fetal loss, endometriosis, or infections in the reproductive tract (*3*). Studies of captive chimpanzees have produced conflicting evidence about the timing or existence of menopause, probably because researchers have used different measures of follicular depletion (e.g., postmortem counts of ovarian sections, observations of sexual swellings) and have included very few individuals in the age range where menopause is expected to occur (*4*–*7*). To assess whether Ngogo female chimpanzees experience menopause, we analyzed variation in the concentration of five hormones diagnostic of menopause. These were measured in 560 urine samples from 66 females that differed in reproductive status and age (range: 14 to 67 years). We compared the patterns found in our sample of chimpanzees with those from previous endocrinological studies of human females.

Results

Post-reproductive representation

The pattern of fertility at Ngogo resembles that in other chimpanzee communities (Fig. 1A). Probabilities of giving birth per year decline after around age 30, and reproduction ends near age 50, as in humans (*8*, *9*). However, whereas females in other chimpanzee communities rarely live past age 50, 16 females at Ngogo have survived past this age (Fig. 1B).

We categorized females as post-reproductive if they were at least 40 years old and lived a long time after their last birth, using the definition provided by Caro and colleagues (*10*). This definition identifies post-reproductive females as those who lived past the age of their last reproduction for longer than the mean plus two standard deviations of successful, closed interbirth intervals for their population. We calculated this value as 7.9 years at Ngogo [5.5 + (2 × 1.2)]. Eleven of the 34 Ngogo females who survived to age 40 (32.4%) were post-reproductive by this criterion (Fig. 2). Nine of these 11 females lived at least 10 years past their age of last reproduction, and, on average, they were observed for 14.1 years without giving birth (SD = 4.6) (Fig. 2). Six of the 11 post-reproductive females lived past the end of our formal observation window in 2016, including three who are still alive now (as of 12 October 2023).

This demographic approach for assaying post-reproductive life spans has several shortcomings (*1*). The criterion of living 7.9 years

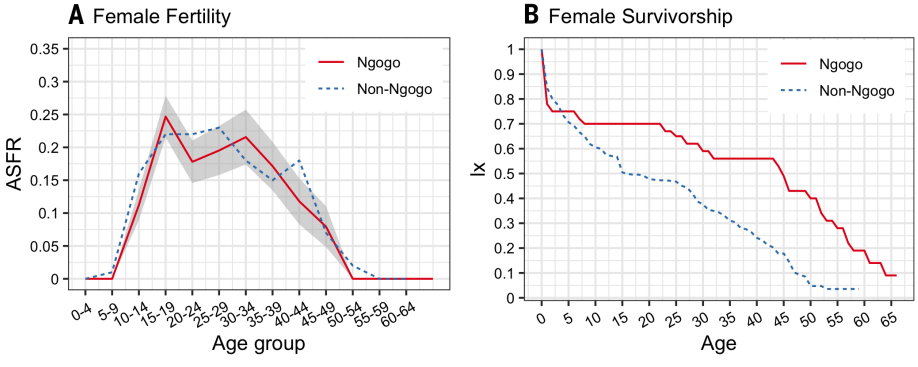


Fig. 1. Ngogo female fertility and survivorship. Female age-specific fertility rates, ASFR, were calculated for each 5-year age group by dividing the total number of births by the years of life observed in females of each age group. (**A**) Ngogo ASFR (±SE, in gray) and a composite sample of six other wild chimpanzee communities reported by Emery Thompson *et al*. (*8*). (**B**) Plot of the probability of female survival to each age (l_x) [updated from Wood *et al*. (*65*)] and a composite sample of five other wild chimpanzee communities (*8*).

[1]Department of Anthropology, University of California, Los Angeles, Los Angeles, CA, USA. [2]Department of Human Behavior, Ecology and Culture, Max Planck Institute of Evolutionary Anthropology, Leipzig, Germany. [3]School of Anthropology, University of Arizona, Tucson, AZ, USA. [4]Center for Species Survival, Smithsonian Conservation Biology Institute, Front Royal, VA, USA. [5]Interim Group Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [6]Comparative BioCognition, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany. [7]Department of Anthropology, University of New Mexico, Albuquerque, NM, USA. [8]Biology Department, McLennan Community College, Waco, TX, USA. [9]Department of Anthropology, Yale University, New Haven, CT, USA. [10]Department of Anthropology, University of Michigan, Ann Arbor, MI, USA. [11]Institute of Human Origins, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA.
*Corresponding author. Email: brianwood@anthro.ucla.edu (B.M.W.); kevin.langergraber@asu.edu (K.E.L.)

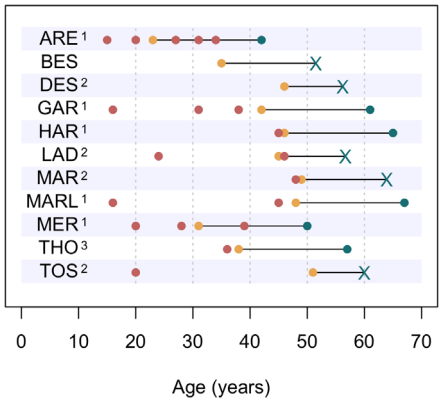


Fig. 2. Ngogo post-reproductive females (1995–2016). Their age at entry to the study is shown by yellow points, and age when giving birth (observed or inferred) is shown by red points. Age when exiting the study is represented by blue points if they were alive at the end of the study period, or by blue crosses if they died during the study period. Numbers plotted next to chimpanzee IDs describe the urine samples collected for hormonal analyses for that individual. Group 1 contributed samples to the analysis of FSH, LH, and ovarian steroid hormones; group 2, only FSH; and group 3, only ovarian steroid hormones. No urine samples were collected from individual BES.

beyond their last birth is likely to miss identifying truly post-reproductive females who died before reaching that milestone (i.e., false negatives). Furthermore, if interbirth intervals increase with age, some "post-reproductive" females may yet reproduce in the future (i.e., false positives), although none of the three living females in the sample have done so at the time of this writing. Finally, even if senescence of the reproductive system typically occurs at the same rate as senescence of the rest of the body, some individuals will deviate from this pattern and stop reproducing well before they die. These considerations indicate the need for alternative methods to determine whether the frequency and duration of post-reproductive life spans are sufficiently large to reject the null hypothesis that somatic and reproductive senescence occur in parallel.

A more informative demographic approach is to calculate the post-reproductive representation (PrR) statistic, which represents the fraction of female adult years lived in a post-reproductive state. This statistic can be validly compared across populations that vary in longevity. The calculation uses the equation

$$\mathrm{PrR} = T_M / T_B$$

where T_M is the expected life years lived beyond age M, T_B is the expected life years lived beyond age B, M is the age at which 95% of lifetime fertility has been realized, B is the age at which 5% lifetime fertility has been realized, and T_M is calculated from life tables using the formula $T_x = l_x \times e_x$, where l_x is the proportion of individuals surviving to exact age x, and e_x is future life expectancy at x.

When survival and fertility decline in parallel, PrR will be near 0. The statistical significance of PrR values greater than 0 can be determined through demographic simulations based on the null hypothesis that declines in survivorship and age-specific fertility occur simultaneously and at the same rate (*1*). Applying these formulas yields values of 14 years for B and 47 years for M at Ngogo, which results in values of T_M = 4.88, T_B = 24.97, and PrR = 0.195. This PrR value of 0.195 indicates that an Ngogo female who reaches adulthood can expect to live about one-fifth her life in a post-reproductive state. This PrR value is significantly different from 0 (see materials and methods), leading us to reject the null hypothesis of parallel declines of survivorship and fertility.

In a study of 52 wild mammal species, including nine nonhuman primates, 49 had PrR values close to 0 (range: 0 to 0.036) and not statistically different from chance under the null hypothesis of parallel declines of survival and fertility (*11*). Similarly, in a study of seven wild nonhuman primate species, including one chimpanzee population with PrR = 0.022, PrR ranged from 0.02 to 0.06 (*12*). Results of prior studies on primates specifically are summarized in Table 1. Only in humans (e.g., Hadza hunter-gatherers, PrR = 0.443), killer whales (0.309), and short-finned pilot whales (0.260) has PrR been shown to be sufficiently large to reject the null hypothesis of concurrent reproductive and somatic senescence under conditions of natural fertility and mortality. Our results are thus the first documentation of a wild nonhuman primate population that exhibits substantial and statistically significant post-reproductive representation.

Hormonal indicators of menopause

In humans, formal diagnosis of menopause typically occurs retrospectively, after a woman has experienced 12 months of amenorrhea; this method of diagnosis is difficult to implement given the observational methods used to study wild chimpanzees. However, menopause also results in a clear endocrinological signature in humans. As follicular stocks near depletion, the ovaries cannot sustain the levels of estrogen and progesterone production necessary for ovulatory cycles. The pituitary gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (FSH) fail to elicit the expected steroid response, and secretion of these hormones consequently increases via negative feedback.

Endocrinological analyses confirm that Ngogo chimpanzees experience menopause (Figs. 3 and 4). As observed in humans, concentrations of ovarian steroid hormones (estradiol, estrone, and pregnanediol, a metabolic product of progesterone) were lower in post-reproductive females (defined as above: ≥7.9 years since last birth at the time of sample collection) than in reproductive females who were regularly experiencing sexual swellings and thus experiencing ovarian cycles at the time of sample collection (2016–2018) (Fig. 3, D to F, and table S1). In another similarity to humans, post-reproductive female chimpanzees showed ovarian steroid hormone concentrations comparable to those of lactating females, indicating the absence of ovarian cycles in both groups (Fig. 3, D to F, and tables S1 and S2).

In humans, the largest change in reproductive hormones across the menopause transition involves gonadotropins. LH and FSH levels spike briefly before ovulation in normally cycling women but are otherwise low, whereas both of these hormones increase and remain elevated from perimenopause into the post-menopausal period. In serum, mean LH levels are two to five times higher and FSH levels are 3 to 15 times higher in post-reproductive women than in reproductive women, although these differences are also influenced by variation in age and breastfeeding status (*13*–*15*). Limited data suggest that the differences may be smaller for urine, with one recent study reporting mean urinary FSH levels approximately two times higher in post-reproductive individuals than in reproductive individuals (*16*). Although methodological differences limit the utility of precise quantitative comparisons, the differences in gonadotropin levels observed between reproductive and post-reproductive Ngogo chimpanzees fell within the range seen in human females. Post-reproductive Ngogo chimpanzees had median LH levels 5.6 times higher and FSH levels 5.4 times higher than those of reproductive females (samples collected from 2016 to 2018) (Fig. 3, A and B, and table S2). In a smaller set of samples collected 10 years earlier using a different assay and including some individuals not available in the later set of samples, FSH levels were 1.8 times higher in post-reproductive than in reproductive individuals (2006–2007 samples) (Figs. 3C and 4C and table S2). Individual-level patterns in the 2016–2018 data (fig. S1) show that all post-reproductive females had FSH and LH values that were higher than the median for reproductive females, except for the post-reproductive female with the earliest age of reproductive cessation (individual ARE, 34 years) (Fig. 2). ARE's reproductive cessation may have therefore been due to a nonmenopausal form of secondary sterility, or she may have been in the early stages of perimenopause, and we underestimated her age (which in our study is generally more likely than age overestimation; see the Materials and methods section).

To verify that reproductive cessation generally occurred as a result of menopause rather than other pathological causes, we also examined the

Table 1. Published estimates of post-reproductive representation for nonhuman primates and selected human populations. The statistical test for PrR is based on simulations of a null model of parallel declines in fertility and mortality, following the procedure described in (*1*). An asterisk in the "Source for demographic measures" column indicates that age-specific measures of mortality and fertility were calculated from a proprietary International Species Inventory System database (*23*) and used to calculate PrR, but user restrictions prevented the authors from publishing or otherwise sharing the underlying data, life tables, or fertility measures.

Species	Community	PrR	PrR source	PrR statistical significance	Source for demographic measures
		Wild nonhuman primates			
Chimpanzee (*Pan troglodytes*)	Ngogo, Kibale, Uganda	0.195	This study	Significant	This study
Chimpanzee (*P. troglodytes*)	Gombe, Tanzania	0.006	(*11*)	Not significant	(*40*)
Chimpanzee (*P. troglodytes*)	Gombe, Tanzania	0.02	(*12*)	Not tested	(*12*)
Chimpanzee (*P. troglodytes*)	Composite of Gombe, Tai, Kanyawara, Mahale, Bossou	0.018	(*29*)	Not significant	(*8*, *66*)
Mountain gorilla (*Gorilla beringei*)	Rwanda	0.022	(*11*)	Not significant	(*11*)
Mountain gorilla (*G. beringei*)	Rwanda	0.04	(*12*)	Not tested	(*12*)
Blue monkey (*Cercopithecus mitis*)	Kakamega Forest, Kenya	0.005	(*11*)	Not significant	(*40*)
Blue monkey (*C. mitis*)	Kakamega Forest, Kenya	0.02	(*12*)	Not tested	(*12*)
Blue monkey (*C. mitis*)	Kakamega Forest, Kenya	0.041	(*29*)	Not significant	(*86*, *87*)
Japanese macaque (*Macaca fuscata*)	Yakushima and Kinkazan Islands, Japan	0.005	(*11*)	Not significant	(*88*)
Rhesus macaque (*Macaca mulatta*)	Community name not listed by (*29*)	0.007	(*29*)	Not significant	No source listed by (*29*)
Northern muriqui (*Brachyteles hypoxanthus*)	Caratinga, Brazil	0.06	(*12*)	Not tested	(*12*)
Olive baboon (*Papio anubis*)	Gombe, Tanzania	0.02	(*11*)	Not significant	(*89*)
Ring-tailed lemur (*Lemur catta*)	Berenty, Madagascar	0.001	(*11*)	Not significant	(*90*)
Verreaux's sifaka (*Propithecus verreauxi*)	Beza Mahafaly, Madagascar	0.003	(*11*)	Not significant	(*40*)
Verreaux's sifaka (*P. verreauxi*)	Beza Mahafaly, Madagascar	0.02	(*12*)	Not tested	(*12*)
White-headed capuchin (*Cebus capucinus*)	Santa Rosa, Costa Rica	0.004	(*11*)	Not significant	(*40*)
White-headed capuchin (*C. capucinus*)	Santa Rosa, Costa Rica	0.04	(*12*)	Not tested	(*12*)
Yellow baboon (*Papio cynocephalus*)	Amboseli, Kenya	0.036	(*11*)	Not significant	(*40*)
Yellow baboon (*P. cynocephalus*)	Amboseli, Kenya	0.01	(*12*)	Not tested	(*12*)
		Captive nonhuman primates			
Chimpanzee (*P. troglodytes*)	Zoo-living composite	0.224	(*29*)	Significant	*
Western gorilla (*Gorilla gorilla*)	Zoo-living composite	0.214	(*1*)	Significant	*
Sumatran orangutan (*Pongo abelii*)	Zoo-living composite	0.231	(*1*)	Significant	*
Bornean orangutan (*Pongo pygmaeus*)	Zoo-living composite	0.192	(*1*)	Significant	*
Japanese macaque (*M. fuscata*)	Provisioned sanctuary, Arashiyama West, Texas	0.054	(*29*)	Significant	(*91*, *92*)
Japanese macaque (*M. fuscata*)	Zoo-living composite	0.247	(*29*)	Significant	*
Rhesus macaque (*M. mulatta*)	Zoo-living composite	0.178	(*29*)	Significant	*
		Humans			
Human (*Homo sapiens*)	Hadza hunter-gatherers, Tanzania	0.443	(*11*)	Significant	(*84*)
Human (*H. sapiens*)	!Kung hunter-gatherers, Botswana	0.426	(*1*)	Significant	(*93*)
Human (*H. sapiens*)	Ache hunter-gatherers, Paraguay	0.439	(*1*)	Significant	(*94*)
Human (*H. sapiens*)	Historical Sweden 1751–1755	0.477	(*1*)	Significant	(*95*)
Human (*H. sapiens*)	Plantation slaves of Trinidad	0.302	(*1*)	Significant	(*60*)

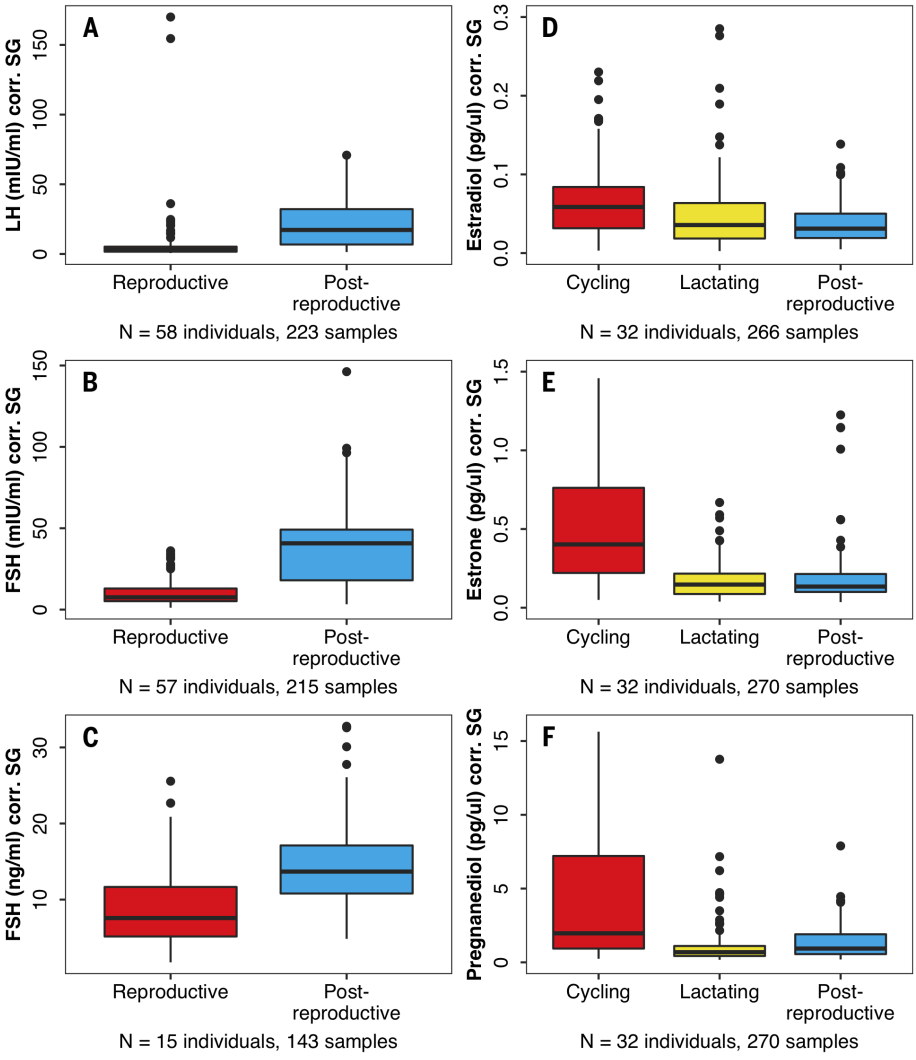


Fig. 3. Urinary hormone concentrations in females of different reproductive states. (**A**) LH in samples collected in 2016–2018, (**B**) FSH 2016–2018, (**C**) FSH 2006–2007, (**D**) estradiol 2016–2018, (**E**) estrone 2016–2018, and (**F**) pregnanediol 2016–2018. Boxplots are based on all urine samples rather than averages per female. Black dots indicate data points above the third quartile by >1.5× the interquartile range. Median values by reproductive category are provided in table S2. For legibility, a few extreme measures have been excluded from (D) to (F), but full data ranges are displayed in fig. S3. corr. SG, corrected specific gravity.

shape of changes in ovarian steroid hormones and gonadotropins with age. If Ngogo female chimpanzees follow a humanlike pattern of FSH changes accompanying menopause at around 50 years of age, there should be modest increases in their FSH levels beginning in their 30s, continuing into their early 40s, and then followed by a more acute increase starting around age 45 (*13*, *17*, *18*). LH concentrations are expected to remain relatively stable before rising in the mid to late 40s (*18*).

Consistent with the human pattern, FSH and LH increased with age (Fig. 4, A to C). To examine the shapes of these relationships, we evaluated alternative regression fits using Akaike's information criteria corrected for small sample sizes (AICc). For the larger 2016–2018 sample, changes in both FSH and LH exhibit steeply increasing nonlinear relationships with age beginning in the mid-40s, rather than linear relationships (Fig. 4, A and B, and table S3). In the 2006–2007 sample, the positive association between age and FSH is best described as a linear relationship (Fig. 4C and table S4), although this result probably stems from a smaller sample size and the inclusion of very few samples from females between the ages of 20 and 50 years. As in humans, gonadotropin production accelerated around the age of typical last reproduction at 50 years.

As Ngogo females age, they produce lower levels of ovarian steroids, especially after the age of 40. Figure 4, D to F, illustrates these age-related decreases in urinary estradiol, estrone, and pregnanediol among nonbreastfeeding Ngogo females. Post-reproductive Ngogo females exhibited uniformly low levels of ovarian hormones compared with reproductive females, a pattern also observed in human females (*19*).

Figure 5 compares standardized measures of gonadotropins and ovarian hormone concentrations in Ngogo females with those of women between the ages of 25 and 60. Because laboratory techniques and units of measure differ across these studies, we have standardized the measures of hormone concentration by their mean values within each study across the age range of 25 to 60, the age range of the human studies. Both species exhibit broadly similar patterns of increasing FSH and LH and declining estrogens and progestins as they undergo menopause.

We recognize that while our samples sizes are large by primatology standards, they are much smaller than those in human studies. Also, our age estimates—particularly those for older females—contain error that is not present in studies of human females (see Materials and methods section "Age estimation"). But even considering these limitations, the similarity across species in the age trends of urinary FSH, LH, estrogens, and progestins is considerable. In sum, the combination of demographic and hormonal data indicates that reproductive cessation in both humans and chimpanzees is caused by a common physiological factor, which is menopause, and that this occurs at a similar age of around 50 years in both species.

Discussion and conclusions

Although the occasional presence of old, nonreproductive females has been reported in other wild chimpanzee communities (*4*, *9*), demographic studies have not previously provided evidence for substantial post-reproductive survival in any wild population. The primary reason for this contrast is that the Ngogo chimpanzees exhibit higher survival rates than those in other populations studied in the wild, where only a few female chimpanzees have been observed to live beyond the age of 50. This demographic difference can be interpreted in two distinct ways, both of which have implications for understanding the evolution of this rare trait in humans.

The first possibility is that the long life spans of the Ngogo chimpanzees are a temporary demographic response to unusually favorable ecological circumstances. In captivity, or so-called "protective" environments, some mammal species, including chimpanzees, exhibit significant female post-fertile reproductive survivorship (Table 1). Leopards were extirpated from Kibale National Park by human hunters in the 1960s, and leopard predation may historically have been an important source of chimpanzee mortality, although this claim is contested (*20*–*22*). While hunters from the high-density human population surrounding the park have occasionally killed Ngogo chimpanzees using snares, metal spears, and trained

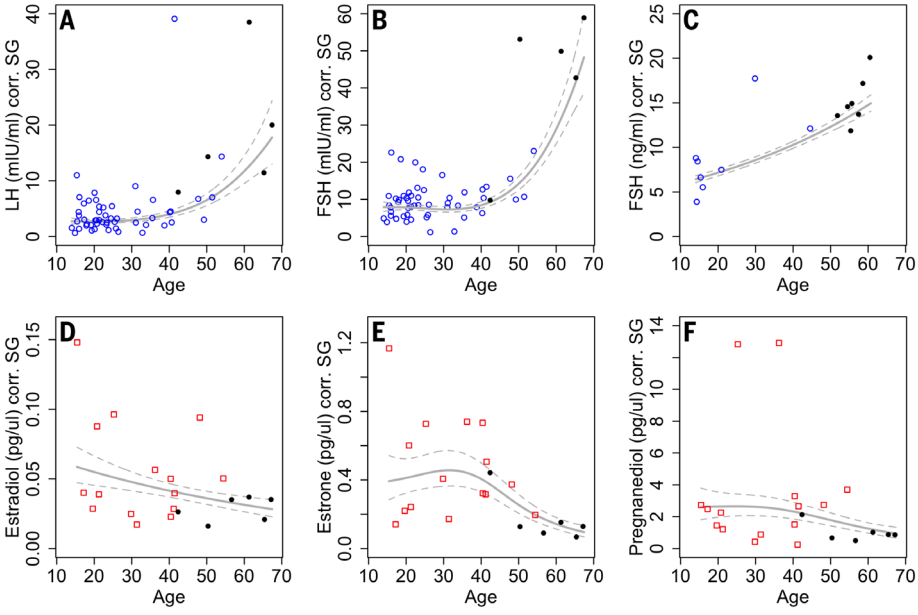


Fig. 4. Urinary concentrations of gonadotropins and ovarian hormones in females by age. (**A**) LH in 223 samples from 58 individuals collected 2016–2018, (**B**) FSH in 215 samples from 57 individuals collected 2016–2018, (**C**) FSH in 143 samples from 15 individuals collected 2006–2007, (**D**) estradiol in 160 samples from 21 individuals, (**E**) estrone in 162 samples in 21 individuals, and (**F**) pregnanediol in 162 samples from 21 individuals. Points represent average levels aggregated by individual female and reproductive status at the time of sample collection. Black points in (A) to (F) represent post-reproductive females, blue points in (A) to (C) represent reproductive females, and red squares in (D) to (F) represent nonbreastfeeding cycling females. The gray age curves in (A) to (F) represent mean expected values (±1 SE) from fit models 10 to 15, respectively (tables S4 to S7). See fig. S3 for sample-level plotting of gonadotropin concentrations and fig. S4 for age trends in ovarian hormone concentrations in all females, including those who were breastfeeding.

packs of dogs, no such cases were known to occur during the study period (*23*). It is thus possible that hunting by humans was a larger source of mortality in chimpanzees' overall long-term evolutionary history than currently occurs at Ngogo. Dietary quality is also high at Ngogo; the local fruit supply is more abundant and stable, and chimpanzees there consume considerably more meat than those in the nearby Kanyawara community (*24*–*26*). The Ngogo chimpanzees have been more successful in between-group competition than chimpanzees in other communities. They expanded their territory by 22% in 2010 after a decade of intense territorial boundary patrolling and killing in the area of the expansion (*27*). Finally, the Ngogo chimpanzee population grew over the study period, with a net reproductive rate (R_0) of 2.0, which suggests that their high survivorship cannot be representative of the long-term evolutionary history of the species.

If the substantial post-reproductive life spans observed in chimpanzees are only a temporary response to favorable and protective ecologies, this would offer insight into the early stages of its evolution as a species-typical trait in humans. Population genetic theory predicts, and empirical research confirms, that the power of selection to modify a trait depends on the amount of its standing genetic variation present in the population (*28*). Substantial post-reproductive life spans cannot evolve as a species-typical trait unless there already exist in the population individuals who outlive their own reproduction (i.e., post-reproductive viability) (*29*). In populations where there are individuals surviving beyond their reproductive years, natural selection can effectively translate the potential indirect fitness benefits of ceasing reproduction into a life stage for the species. The capacity for this demographic response in the last common ancestor humans share with chimpanzees would thus provide modest support for the plausibility of substantial post-reproductive life spans evolving earlier in hominin evolution [around 1.8 million years ago, at the emergence of *Homo erectus*, as discussed in (*30*)] rather than only very recently (around 50 thousand years ago, at the dispersal of *Homo sapiens* from Africa).

The second possibility is that substantial post-reproductive survivorship has been common in the evolutionary history of chimpanzees, but it is not exhibited by contemporary populations elsewhere because of recent environmental changes caused by humans. With one exception [Tai Forest (*31*)], previous demographic studies of chimpanzees have been conducted on populations living in habitats that were more heavily logged for forestry or agriculture within the past 100 years than Ngogo and are thus composed of less primary, old-growth forest (*32*–*35*). In addition, their close evolutionary relationship to humans makes chimpanzees extremely vulnerable to respiratory viruses that originate in humans and to which they have little developmentally acquired adaptive immunity or evolutionarily acquired genetic immunity [see review in (*36*)]. The devastating impact of human diseases was made clear by an outbreak of human metapneumovirus at Ngogo that killed 12.2% of the individuals in the community in early 2017, after the period under study here (*37*). The Ngogo chimpanzees had not previously suffered any known major disease outbreaks, and they have lower viral richness and load than the Kanyawara chimpanzee community, which lives at the edge of the park and has more contact with humans (*38*). Although more systematic research is needed, one study (*39*) concluded that anthropogenically caused habitat loss and disease epidemics are likely the main drivers of the substantial variability in survival rates observed across chimpanzee populations and for the fact that most have recently experienced devastating declines. The negative growth rates of non-Ngogo chimpanzee populations, like the positive growth rate of Ngogo, indicate that their patterns of survivorship and fertility also cannot represent the long-term average for the species. Comparative data suggest that favorable ecological conditions promoting population growth do not alone account for the emergence of substantial post-reproductive representation, as six other species of wild primates studied by Bronikowski and colleagues (*40*) have also grown at robust rates similar to those of the Ngogo chimpanzees (R_0 range: 1.51 to 2.37) yet do not show substantial post-reproductive life spans.

If substantial post-reproductive life spans were more common in the evolutionary history of chimpanzees, this would have additional implications for evaluating arguments for the evolution of prolonged post-reproductive life spans. A prominent adaptive evolutionary explanation for the evolution of substantial post-reproductive life spans in humans is the grandmother hypothesis (*41*, *42*), some versions of which state that the indirect fitness benefits that post-reproductive females gain by helping their daughters to reproduce or their grandoffspring to survive are greater than the direct fitness costs of ceasing reproduction. However, whether such indirect fitness benefits are sufficiently large to outweigh the direct fitness costs of ceasing direct reproduction in humans is controversial (*43*, *44*). Such indirect fitness benefits would likely be even smaller in chimpanzee-like social contexts. Chimpanzee males typically remain in the communities in which they were born for their entire lives. By contrast, females typically disperse from their natal community to reproduce

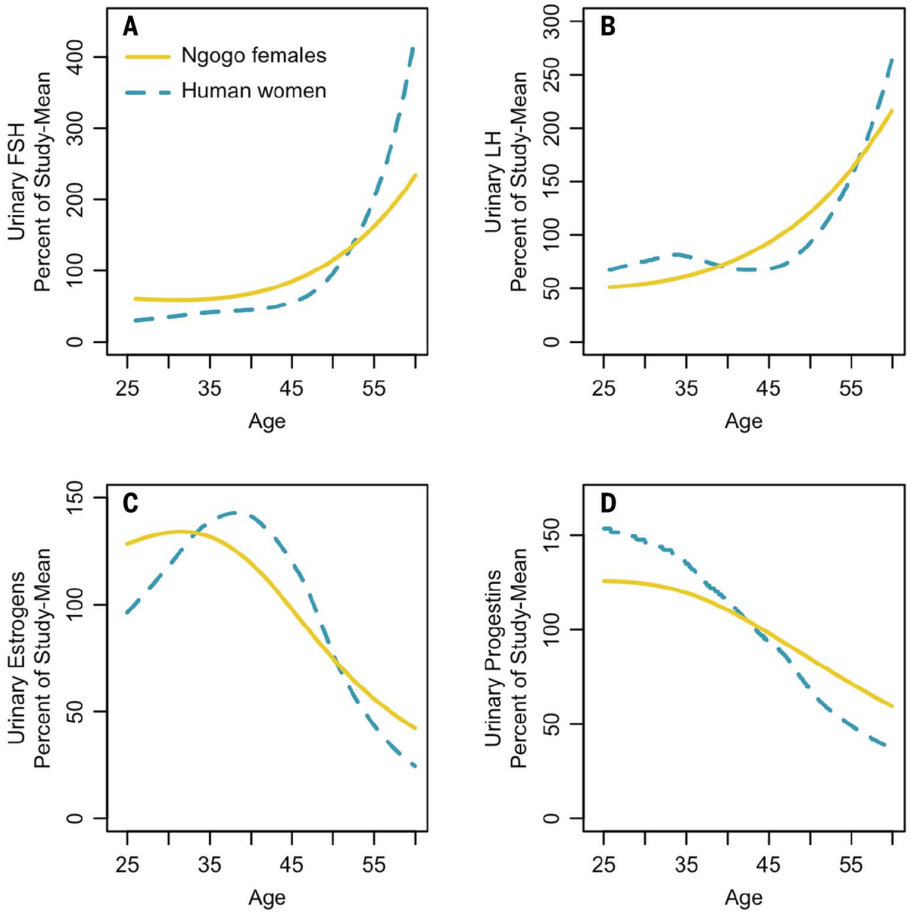


Fig. 5. Standardized age trends in female urinary hormones in Ngogo chimpanzees and human females. (**A**) FSH, (**B**) LH, (**C**) estrogens, and (**D**) progestins, as observed at Ngogo and in studies of human females by Ferrell *et al*. (*18*, *19*). The expected values for Ngogo females in all panels were generated from our study's data and the fit models 10, 11, 14, and 15 (tables S4 and S6). The analytes from Ngogo females shown in (C) and (D) are estrone and pregnanediol, respectively. Median age trends for human females in urinary FSH, LH, estrone glucuronide (a conjugated metabolite of estrone), and pregnanediol glucuronide (a conjugated metabolite of progesterone) are derived from (*18*, *19*).

in others, where they remain until they die (*45*). Unlike humans, chimpanzees cannot maintain long-distance helping relationships with individuals who reside in different communities (*46*, *47*). Thus, female dispersal limits opportunities for chimpanzee grandmothers to help either their daughters or their daughters' offspring, who will generally live elsewhere. Post-reproductive chimpanzee females could potentially help their sons reproduce, but current evidence suggests that unlike female bonobos, they do not do so. The difficulties of recognizing paternal kin under a highly promiscuous, polygynadrous mating system probably limits the ability of females to help their sons' offspring survive (*48*–*51*).

Another prominent adaptive evolutionary explanation for substantial post-reproductive life spans is the reproductive conflict hypothesis (*52*). This hypothesis suggests that in certain contexts, older females will compete with younger females for limited reproductive opportunities. In dispersal and mating systems where female relatedness to other breeders and their offspring increases with age, the net benefits of continued reproduction at advanced ages are low, because reproduction imposes costs on relatives. This pattern of age-graded local relatedness occurs under the system of male philopatry and female dispersal that characterizes chimpanzees and, more controversially, may be ancestral in humans and their hominin relatives (*52*–*56*). Some of the assumptions of the reproductive conflict model have been supported in empirical tests with killer whales (*57*), but similar investigations have led to mixed results in humans (*58*) and long-finned pilot whales (*59*).

The grandmother hypothesis and the reproductive conflict hypothesis are not mutually exclusive alternatives, and both may be required to explain why humans have evolved a robust post-fertile life stage. Even in the harshest socio-ecological conditions, human communities still have a higher representation of post-reproductive individuals compared with Ngogo chimpanzees. For example, under the extreme conditions of plantation era slavery in Trinidad, historical demographers estimate that a PrR of 0.302 was maintained (*1*, *60*), which is about 1.5 times that seen at Ngogo.

In sum, while the evolution of menopause and a substantial post-reproductive life span remains unclear, our results show that these traits can emerge in a chimpanzee population that has experienced a low level of human impact. The capacity for long post-reproductive life spans observed in contemporary humans may not have evolved de novo in our hominin ancestors but instead built on existing genetic variation in the last common ancestor we shared with chimpanzees. Our documentation of the survival of post-reproductive females at Ngogo required an extensive and ongoing research effort, and it will be crucial to invest in long-term studies across diverse ecological settings to better understand whether substantial post-reproductive life spans in chimpanzees emerge only rarely in particularly favorable ecological circumstances or are actually more common. There are currently insufficient data on survival and fertility in chimpanzees' sister species bonobos (*Pan paniscus*) to ascertain whether substantial post-reproductive life spans also occur in this species, and if so, under what conditions. Filling these gaps in knowledge about post-reproductive life spans in humans' two closest living relatives will be important for our understanding of the evolution of this rare and puzzling trait.

Materials and methods

Study site

All demographic and endocrinological data from females came from the Ngogo community of chimpanzees. Ngogo is located in Kibale National Park, southwestern Uganda. The 795 km^2 park is centered at ~0.5°S and 30.4°E. The vegetation of the park is mostly moist evergreen or semi-deciduous forest, transitional between lowland and montane forest (*33*). The Ngogo study area is in the center of the park, at altitudes between 1400 and 1470 m. Ngogo receives ~1500 mm of rainfall annually, concentrated in March to May and September to December (*61*, *62*). The vegetation is typically dry-ground forest that includes large tracts of old growth forest and early- to mid-stage colonizing forest regenerating from anthropogenic grassland (*63*). Anthropogenic grasslands still cover some of the study area, which also includes swamp forest, bush dominated by *Acanthus pubescens*, and a papyrus (*Cyperus papyrus*) swamp (*64*). The chimpanzees at Ngogo use the old-growth forest predominately and stay entirely within the park. Ngogo is surrounded on all sides by other communities of chimpanzees and does not

abut areas occupied by humans. Consequently, the chimpanzees do not eat crops but instead rely entirely on wild foods. Veterinary interventions have been limited to the removal of a snare from one adult male.

Demographic data

Our prior study of mortality (*65*) provides in-depth descriptions of our methods of demographic data collection, age assessment, and parentage assignment. Information on Ngogo chimpanzee births, interbirth intervals, age-specific fertility rates (ASFR), and post-reproductive survivorship are reported here for the first time, using data that span the years 1995–2016, including 306 individual chimpanzees and 3108 chimpanzee-risk-years of observation. Owing to the large size of the Ngogo community, the Ngogo demographic database is among the largest available for any wild chimpanzee community. This sample includes 185 females and 1611 risk years of observation.

Because the Ngogo chimpanzees have only been observed continuously since 1995, the ages of the oldest chimpanzees were necessarily estimated through morphological comparisons with known-aged individuals and genetically determined kinship relationships [as in (*65*, *66*) and see Materials and methods section "Age estimation"]. However, our demographic and physiological data on reproduction provide a strong internal validation of these age estimates. First, ASFR at Ngogo closely accord with data from other chimpanzee populations (Fig. 1). Second, the ages of post-reproductive females and ages at which FSH and LH levels increased exceeded the maximum ages of the majority of chimpanzees previously described from the wild, suggesting why long post-reproductive life spans were found at Ngogo but not elsewhere (Fig. 4). As a robustness check, we created a simulation model that generates alternative schedules of mortality and fertility, on the basis of a reasonable model of age estimation error. Including these influences of age estimation error, Ngogo post-reproductive representation (PrR) remains substantial and significant (Materials and methods section "The statistical significance of post-reproductive representation is robust to error in age estimates"). When we subsample our data and focus only on individuals who were precisely aged, we also find that Ngogo survivorship is high, consistent with our general results (Materials and methods section "Precisely aged individuals show low mortality among the Ngogo chimpanzees").

Interbirth intervals

Seventy-four females gave birth at Ngogo during the study, to a total of 175 offspring. We also identified 37 additional mother-offspring dyads through genetic analyses and observations of especially close interactions between females and young. Births that occurred before our observation period were not included in our analyses of ASFR or interbirth intervals, but they did allow us to infer whether some females had ever given birth. We identified closed, successful interbirth intervals at Ngogo—that is, those closed interbirth intervals in which first offspring survived to at least 4 years of age (n = 75). The mean duration of these intervals was 5.5 years (SEM = 0.14) and the standard deviation was 1.2 years.

Demographic measure of post-reproductive life span

We sought to identify older females who had not reproduced for an extended period and who were thus strong candidates for having experienced menopause (Fig. 1). We defined "older" as estimated to be at least 40 years and "extended period" as the population mean interbirth interval plus two standard deviations (7.9 years).

The PrR statistic is constructed from the observed life tables of survivorship and fertility and should be near zero for species in which a substantial female post-reproductive life span does not occur. However, owing to sampling variation, a nonzero PrR value is insufficient evidence to conclude that post-reproductive representation is generally greater than zero in a study population. Levitis and Lackey (*1*) present a demographic simulation procedure that generates PrR values from the null hypothesis that age-specific fertility rates (m_x) and survivorship (l_x) decline in parallel after the age of peak fertility. The observed age-specific fertility rates (m_x) are used to construct a smoothed model of monotonically declining fertility after the age of peak fertility. This fertility decline model is then used to construct the predicted survivorship $(\widehat{l_x})$ that would be anticipated if the null hypothesis were true, that is, if declines in survivorship were in parallel with and directly proportional to declines in fertility. Using the observed m_x and null-predicted survivorship values $(\widehat{l_x})$, 1000 simulated populations are then created, each with its own corresponding null-predicted $\widehat{\mathrm{PrR}}$ statistic. The distribution of these null $\widehat{\mathrm{PrR}}$ values is then compared with the observed PrR. The P value for this significance test is the number of null-simulated populations with higher $\widehat{\mathrm{PrR}}$ values than the observed PrR, divided by 1000. In our case, no null-simulated populations had a $\widehat{\mathrm{PrR}}$ value higher than 0.196; thus, our P value is <0.001.

Collection and measurement of urinary reproductive hormones

Most urine samples were collected opportunistically from individually identified female chimpanzees between 1 March 2016 and 11 May 2018. A second smaller dataset of urine samples from which we only analyzed FSH was collected between 28 May 2006 and 25 November 2007. Urine was pipetted from plastic sheets or leaves (*67*) and transferred to collection vials. Vials were stored in a cooled thermos until the collector returned to camp then deposited in a solar-powered freezer at −20°C. Samples designated for steroid measurements were transported on dry ice from Uganda to Germany, where they were stored at −80°C until they were analyzed. Similarly, samples designated for gonadotropin measurements were transported on ice from Uganda to the United States and stored at −80°C.

Urine samples collected in 2016–2018 were analyzed for gonadotropins FSH and LH in the Comparative Human and Primate Physiology Center at the University of New Mexico, Albuquerque, NM, USA. FSH and LH data from 2016–2018 were assayed using commercial enzyme-linked immunosorbent assays from MP Biomedicals with a sensitivity of 1.5 mIU/ml FSH and 1.0 mIU/ml LH. Interassay coefficients of variation (CVs) were 9.1% (low control) and 8.6% (high) for the FSH assay and 6.7% (low) and 13.4% (high) for the LH assay. Intrassay CVs, calculated as the mean CV of duplicate determinations, averaged 10.1% for both assays.

Urine samples from 2006–2007 were analyzed for FSH at the Smithsonian Conservation Biology Institute, Front Royal, VA, USA. FSH (LER-1976A) was iodinated following the protocol described in (*68*). Urine FSH was quantified by a iodine-125 double-antibody radioimmunoassay (RIA), with some modifications. The RIA used an anti-ovine FSH primary antibody (JADLER 178) and ovine LH label and standards (NIDDK-FSH-S16) in a phosphate-buffered saline (PBS)–based (0.01 M PO_4, 0.9% NaCl, 0.5% bovine serum albumin, 2 mM EDTA, 0.01% thimerosal, pH 7.4) buffer system. It was incubated in a total volume of 500 μl in 12 mm by 75 mm borosilicate tubes at room temperature. Standards (100 μl) and/or sample were added to PBS (200 μl), followed by addition of primary antibody (1:25,000, 100 μl) and incubation for 24 hours. Iodine-125 FSH tracer (~25,000 counts per minute, 100 μl) was added and the tubes incubated for an additional 24 hours. Antibody-bound complexes were precipitated by centrifugating at 3000g for 25 min after a 1 hour incubation with goat anti-mouse gamma globulin (1:300, 1 ml in PBS containing 5% polyethylene glycol, 8000 molecular weight, Sigma Chemical Co., St. Louis, MO). The antibody typically bound 30% of the iodinated tracer with ~5% nonspecific binding. Assay sensitivity was 0.25 ng/ml. The assay was validated for chimpanzee urine by demonstrating (i) parallelism between dilutions of pooled urine and the standard curve and (ii) significant recovery (>98%) of FSH standard added to non-estrous chimpanzee urine (y = 0.912x + 4.998, r = 0.993).

Urine samples collected in 2016–2018 were analyzed for steroid hormones in the Department

of Primatology at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Each sample underwent extraction before measurement using a modification of Hauser *et al.* (*69*) in which we combined each sample with an internal standard, performed hydrolysis with β-glucuronidase derived from *Escherichia coli*, and performed solvolysis with ethyl acetate and sulfuric acid. Estradiol, estrone, and pregnanediol were measured by liquid chromatography–tandem mass spectrometry (LC-MS) using a modification of (*69*). We used a Waters ACQUITY UPLC separation module for liquid chromatography, using water (with 0.1% formic acid) and acetonitrile as eluents. We then used a Xevo TQ-S tandem quadrupole mass spectrometer (Waters, Milford, MA, USA) for mass spectrometry and examined the output with MassLynx (QuanLynx-Software).

We corrected for variation in urine concentration by measuring the specific gravity (SG) of each sample. We measured SG with a digital handheld refractometer (2016–2018 samples for gonadotropins: PAL-10S, ATAGO, Bellevue, WA, USA; 2006–2007 samples for FSH: Urisystem, Fisher Scientific, Waltham, MA, USA; 2016–2018 samples for steroids: TEC, Ober-Ramstadt, Germany). Hormone values were adjusted for SG following (*70*). We present estradiol, estrone, and pregnanediol concentrations in picograms per microliter of SG, and FSH and LH concentrations from 2016–2018 samples in milli–international units per milliliter of SG.

Of the eight reproductive females in the 2006–2007 FSH dataset, six were also included among the reproductive females in the 2016–2018 FSH dataset. Of the nine post-reproductive females in the 2006–2007 FSH dataset, four were also among the post-reproductive females in the 2016–2018 FSH dataset.

All analyses were based on individuals ≥14 years at the time of sample collection, which is close to the average age of first pregnancy among natal female chimpanzees (*71*). The oldest individual was estimated at 67 years. All analyses were based on samples from females that were known (via negative pregnancy tests) or inferred not to be pregnant because their next births occurred more than 253 days after sample collection, which corresponds to the mean gestation length plus two standard deviations in a sample of 118 captive chimpanzees (*72*). Finally, we did not include samples from females who died within 253 days of sample collection and did not have offspring younger than 3 years old, as they might have been pregnant.

Analysis of hormone levels by reproductive category

We assigned females to one of three reproductive categories: (i) post-reproductive, (ii) cycling (i.e., females who were experiencing regular sexual swellings), or (iii) lactating. To compare hormone levels among females in these categories, we used linear mixed models (LMMs) with Gaussian error structures and fitted with restricted maximum likelihood using the "lmer" function in package lme4 (*73*) in R version 4.0.3. For the analysis of gonadotropins (LH and FSH), cycling and lactating females were lumped into a single "reproductive" category whose levels were then compared with post-reproductive females (table S2, models 1 to 3). For the analysis of the steroid hormones estradiol, estrone, and pregnanediol, hormone levels were compared among all three reproductive categories, using 2016–2018 samples (table S2, models 4 to 6). To control for multiple sampling of individuals, we included subject ID as a random intercept in all models. Sexual swellings in chimpanzees coincide with the periovulatory period when sex-steroid and gonadotropin levels are strongly elevated. To compare baseline levels of the different reproductive categories, we omitted samples of females when they exhibited maximal sexual swellings, thereby making statistical comparisons considerably more conservative.

To produce conservative probability values (*74*), we generated Satterthwaite approximations for degrees of freedom using package *lmerTest* (*75*). We set alpha to 0.05 and used the "confint" function to calculate bootstrapped 95% confidence intervals of fixed effects from 1000 simulations. The regression assumption of residual normality was largely satisfied by Box-Cox transforming hormone values before analysis (*76*, *77*). We visually inspected residual plots and qq-plots (*78*) of models 1 to 6 (table S3) fit to all the hormone datasets to assess the normality of residuals and the presence of outliers. This assessment indicated that there were a few extreme observations that fell well outside the model-predicted ranges of hormone values. We then applied a conservative procedure described in Kutner *et al.* (*79*) to identify and exclude outliers from the datasets used in the final analyses. We first calculated the semi-studentized residuals of each observation in the raw data with respect to models 1 to 6. We then excluded any data in which the absolute value of the semi-studentized residual was greater than or equal to 4. This process led us to exclude one sample from the estradiol data, one sample from the estrone data, and one sample from the pregnanediol dataset. No outliers were excluded from the LH or FSH data. The sample sizes in the final hormone datasets are tallied in table S2. The structure of the statistical models (models 1 to 6) used in the analysis of hormone concentrations by reproductive category are listed in table S3.

Analysis of LH and FSH by age

We conducted generalized additive modeling to estimate linear and nonlinear continuous relationships between age and gonadotropin levels (table S4, models 7 to 12). To adjust for repeated samples of individuals, subject ID was treated as a random intercept in all models. Models were fit using the "gam" function of the R package *mgcv* (*80*). The *mgcv* package estimates relationships between predictor and outcome variables using penalized regression splines in which smoother relationships are preferred to more complex or "wigglier" relationships to prevent over-fitting.

In models 7 to 12, the hormone data were log-transformed before analysis, owing to the non-normal distribution and strictly positive ranges of these measures. In models 7 to 9, we specified that the relationship between age and hormone measures must obey a strictly linear relationship, whereas in models 10 to 12, we specified that a nonlinear relationship, constructed using a cubic spline, could characterize the age relationship. We then compared the estimated out-of-sample predictive accuracy of the linear and nonlinear models fit with each sample using AICc.

Comparison of models 7 to 9 to models 10 to 12 shows that treating age as a smoothed predictor variable leads to lower AICc values, representing improved out-of-sample predictive accuracy, for models 10 and 11. In the case of model 12, the penalized regression estimation process in *mgcv* shrunk the spline to a perfectly linear relationship between age and FSH, equivalent to the linear model 9, with an equivalent AICc value. AICc values are reported in table S4, and model coefficients are reported in table S5.

Analysis of estradiol, estrone, and pregnanediol concentrations by age

To investigate the continuous relationship between age and steroid hormone levels, we fit six generalized additive mixed models (models 13 to 18) in which the relationship between age and the hormone value was estimated as a smooth cubic spline using penalized regression estimation. Subject ID was entered into the model as a random intercept to model the possibility of subject-specific differences in mean hormone measures. The structure of these models is reported in table S6, and the coefficients estimated from models 13 to 18 are summarized in table S7. The model-predicted relationships between age and average steroid hormone measures are displayed in Figs. 4 and 5 and figs. S3 and S4.

Comparison of age trends in urinary hormone concentrations in human females and Ngogo females

We calculated average Ngogo female age trends for urinary FSH using model 11 (Fig. 5A), for LH using model 10 (Fig. 5B), for estrone using model 14 (Fig. 5C), and for pregnanediol using model 15 (Fig. 5D). The median age trends for human females for urinary FSH were carefully traced from figure 1B of (*18*), for LH figure 1A

of (*18*), for estrone-glucuronide figure 2A of (*19*), and for pregnanediol-glucuronide figure 2B of (*19*) using WebPlotDigitizer (*81*). Estrone-glucuronide is a metabolite of estrone that humans and chimpanzees excrete in their urine (*82*). In our analyses, following the protocol of (*69*), estrone metabolites were deconjugated to estrone before measurement by LC-MS, whereas in (*19*), immunoreactive estrone-glucuronide was measured directly. Likewise, the major urinary metabolite of progesterone, pregnanediol-glucuronide, was deconjugated to pregnanediol before measurement using LC-MS in our study, whereas immunoreactive pregnanediol was measured directly in (*19*).

There are several data limitations to Fig. 5. First, the error in chimpanzee age estimation would be expected to flatten out steep changes in age trends that might otherwise be detected with more-precise age estimates. Second, Fig. 5 displays only model-expected central trends for each population, ignoring considerable between-individual variability in base rates and change over time that is known to exist from studies of human females [see figure 2 of (*18*)]. Finally, our sample sizes are large by primatology standards but much smaller than those of the human studies.

Age estimation

Ages of the oldest females in the sample, including those found to be post-reproductive, were necessarily estimated because they were adults when first identified. Nevertheless, these estimates are predicated on critical data that together suggest that any error in age estimation does not affect our conclusions.

First, age estimations for females in the study were made when females were first identified, rather than at the time of the current study. For the oldest females in the sample, including those found to be post-reproductive, these determinations were made in 1995–1998, 11 to 23 years before sampling for this study. In other words, it was not the case that females in the study were visually estimated to be >50 years old, but rather they were identified when still young or middle-aged and subsequently observed over many years. Age estimations at the outset of the study benefitted from understandings of visual signs of aging (e.g., gray hair, prominent hips and shoulders, worn or missing teeth) generated by previous chimpanzee research [e.g., long-term research at Gombe (*83*)], and the distribution of these original age estimations for chimpanzees in the Ngogo community were consistent with population structures observed for wild chimpanzees elsewhere. Fortunately, females disperse into new communities within a narrow age range during adolescence, which means that those females who joined the Ngogo community over the course of the study could be aged with minimal error.

In most cases, contextual information allows us to identify the minimum age of females when first observed, on the basis of the presence and size of dependent offspring and/or genetic pedigree data that can identify adult offspring. For example, a female observed with two dependent offspring when first observed should be at least 20 years old. For 4 of the 11 post-reproductive females, they must have already been in their 30s to 40s when first identified in 1995 because they already had genetically identified adult-aged offspring at this time. These estimates will tend to be conservative because first-born offspring may have died or emigrated before the period of genetic sampling. As an illustrative example, consider the oldest post-reproductive female in our sample, MARL. We identified DO as a middle-aged male in 1995 and assigned him the age of 28 on the basis of his physical appearance, making his birth year 1967. Genetic parentage analyses subsequently confirmed our suspicion that adult female MARL was DO's mother. On average, immigrant female chimpanzees have their first offspring at age 16.2 years, which would make MARL's birth year 1951 if we assume DO was her first offspring. However, we instead assigned MARL a birth year of 1950 (and thus age of 45 years in 1995) instead of 1951 in recognition of the possibility that DO was not her first offspring, and because she looked a bit older than females whose age estimates were similarly genealogically informed in 1995. Our age estimate of 28 years for DO in 1995 could not have been considerably overestimated, because when he died in 2014 at an estimated age of 47 years, he had lost considerable muscle mass and all of his canines. Similarly, when MARL died in 2019 at the estimated age of 69, she was very thin, had sparse gray hair, very worn teeth, and had multiple great-grandoffspring.

As noted above, although our age estimates were derived long before this study was conceived, the results help reinforce the validity of the age estimates because ASFR at Ngogo align closely with patterns observed in other chimpanzee populations (Fig. 1), and ages of the menopausal transition conform to the expectations from prior studies of ASFR and follicular depletion rate (Fig. 4).

The statistical significance of post-reproductive representation is robust to error in age estimates

To assess the possible impacts of age estimation error on our measure of post-reproductive representation (PrR) at Ngogo, we created a simulation model that generates alternative schedules of mortality and fertility, embodying a reasonable model of age measurement error. Our model of age estimation error starts with the intuition of human demographers and primatologists that as the actual age of an individual increases, the error in age estimations also increases (*65*, *84*). We treated age measurement error as a process arising from uniform distributions, with ranges that increased proportional to our best point estimate (BPE) of an individual's age when they entered our study. Following (*65*), we used the formula below to specify the relationship between the range of the uniform distribution and our best point estimates

$$\text{Range of uniform distribution} = \text{BPE} * 0.2$$

Applying this model of age estimation error, we generated 1000 data permutations. Among permutations, the simulated ages were distributed across intervals with a minimum value of BPE − BPE∗0.1 and a maximum value of BPE + BPE∗0.1. The scaling factor of 0.2 was chosen because it produces estimates that accord with experience. For example, a chimpanzee with a real age of 10 years might have been visually estimated as being between 9 and 11 years old, while chimpanzees estimated to be ages 20 or 40 when first identified were treated in this model as having probabilistic ages uniformly distributed across the intervals 18–22 and 36–44 years, respectively.

By simulating 1000 permutations of the demographic data, we created 1000 schedules of mortality and fertility. Using these schedules, we then calculated the PrR statistic for each permutation, which resulted in the distribution of values shown in fig. S5.

Among all 1000 data permutations, the minimum PrR value recorded was 0.132. We then applied the analysis procedure described in (*1*) to assess whether even this minimal-PrR value of 0.132 was nevertheless greater than expected under the null hypothesis of parallel declines in survivorship (l_x) and fertility (m_x). We found that the PrR value of 0.132 is indeed higher than expected under the null hypothesis of there being no post-reproductive representation in this community ($P < 0.001$). In summary, while our age estimates may be off by a few years, even when making allowances for such error, Ngogo PrR values are significantly higher than expected if survivorship and fertility declined in parallel in this study population.

Precisely aged individuals show low mortality among the Ngogo chimpanzees

When we subsample our data and only focus on those individuals who were two years or younger when they first entered our mortality risk pool, we see that Ngogo survivorship to young adulthood is higher than in the nearby community of Kanyawara, also located in Kibale National Forest [see figure 2 in (*65*)]. This analysis is particularly informative because age estimation error is minimized in this sample, and comparative demographic analysis indicates that high survivorship in early life

is positively associated with high survivorship across the life course (*85*).

In summary, when making reasonable data adjustments for age estimation error (fig. S6), and when restricting analyses to only those individuals with the highest age accuracy, we find that survivorship among the Ngogo chimpanzees is high and post-reproductive representation is significant. Both age measurement error and sampling variation are thus not adequate alternative explanations for our results.

REFERENCES AND NOTES

1. D. A. Levitis, L. B. Lackey, A measure for describing and comparing post-reproductive lifespan as a population trait. *Methods Ecol. Evol.* **2**, 446–453 (2011). doi: 10.1111/j.2041-210X.2011.00095.x; pmid: 22081792
2. R. E. Jones, K. H. Lopez, *Human Reproductive Biology* (Academic Press, ed. 4, 2013).
3. M. Emery Thompson, K. Sabbi, "Evolutionary demography of the great apes" in *Human Evolutionary Demography*, O. Burger, R. Lee, and R. Sear, Eds. (Open Science Foundation, 2019); https://osf.io/d2thj/.
4. J. G. Herndon *et al*., Menopause occurs late in life in the captive chimpanzee (*Pan troglodytes*). *Age (Dordr.)* **34**, 1145–1156 (2012). doi: 10.1007/s11357-011-9351-0; pmid: 22189910
5. C. E. Graham, Reproductive function in aged female chimpanzees. *Am. J. Phys. Anthropol.* **50**, 291–300 (1979). doi: 10.1002/ajpa.1330500302; pmid: 218459
6. E. N. Videan, J. Fritz, C. B. Heward, J. Murphy, The effects of aging on hormone and reproductive cycles in female chimpanzees (*Pan troglodytes*). *Comp. Med.* **56**, 291–299 (2006). pmid: 16941957
7. A. Lacreuse *et al*., Menstrual cycles continue into advanced old age in the common chimpanzee (*Pan troglodytes*). *Biol. Reprod.* **79**, 407–412 (2008). doi: 10.1095/biolreprod.108.068494; pmid: 18495682
8. M. Emery Thompson *et al*., Aging and fertility patterns in wild chimpanzees provide insights into the evolution of menopause. *Curr. Biol.* **17**, 2150–2156 (2007). doi: 10.1016/j.cub.2007.11.033; pmid: 18083515
9. J. H. Jones, M. L. Wilson, C. Murray, A. Pusey, Phenotypic quality influences fertility in Gombe chimpanzees. *J. Anim. Ecol.* **79**, 1262–1269 (2010). doi: 10.1111/j.1365-2656.2010.01687.x; pmid: 20412347
10. T. M. Caro *et al*., Termination of reproduction in nonhuman and human female primates. *Int. J. Primatol.* **16**, 205–220 (1995). doi: 10.1007/BF02735478
11. S. Ellis *et al*., Postreproductive lifespans are rare in mammals. *Ecol. Evol.* **8**, 2482–2494 (2018). doi: 10.1002/ece3.3856; pmid: 29531669
12. S. C. Alberts *et al*., Reproductive aging patterns in primates reveal that humans are distinct. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13440–13445 (2013). doi: 10.1073/pnas.1311857110; pmid: 23898189
13. H. G. Burger *et al*., Prospectively measured levels of serum follicle-stimulating hormone, estradiol, and the dimeric inhibins during the menopausal transition in a population-based cohort of women. *J. Clin. Endocrinol. Metab.* **84**, 4025–4030 (1999). pmid: 10566644
14. S. J. Lee, E. A. Lenton, L. Sexton, I. D. Cooke, The effect of age on the cyclical patterns of plasma LH, FSH, oestradiol and progesterone in women with regular menstrual cycles. *Hum. Reprod.* **3**, 851–855 (1988). doi: 10.1093/oxfordjournals.humrep.a136796; pmid: 3141454
15. I. Overlie, M. H. Moen, L. Morkrid, J. S. Skjaeraasen, A. Holte, The endocrine transition around menopause—A five years prospective study with profiles of gonadotropines, estrogens, androgens and SHBG among healthy women. *Acta Obstet. Gynecol. Scand.* **78**, 642–647 (1999). pmid: 10422913
16. Y. Onizuka *et al*., Association between FSH, E1, and E2 levels in urine and serum in premenopausal and postmenopausal women. *Clin. Biochem.* **73**, 105–108 (2019). doi: 10.1016/j.clinbiochem.2019.08.009; pmid: 31442440
17. M. R. Sowers *et al*., Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. *J. Clin. Endocrinol. Metab.* **93**, 3478–3483 (2008). doi: 10.1210/jc.2008-0567; pmid: 18593767
18. R. J. Ferrell *et al*., Monitoring reproductive aging in a 5-year prospective study: Aggregate and individual changes in luteinizing hormone and follicle-stimulating hormone with age. *Menopause* **14**, 29–37 (2007). doi: 10.1097/01.gme.0000227859.50473.20; pmid: 17019379
19. R. J. Ferrell *et al*., Monitoring reproductive aging in a 5-year prospective study: Aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. *Menopause* **12**, 567–577 (2005). doi: 10.1097/01.gme.0000172265.40196.86; pmid: 16145311
20. C. Boesch, *The Real Chimpanzee: Sex Strategies in the Forest* (Cambridge Univ. Press, 2009).
21. M. N. Muller, Review of *The Real Chimpanzee: Sex Strategies in the Forest*, by C. Boesch (Cambridge University Press, 2009). *Int. J. Primatol.* **32**, 524–529 (2011). doi: 10.1007/s10764-010-9481-8
22. K. Zuberbühler, D. Jenny, Leopard predation and primate evolution. *J. Hum. Evol.* **43**, 873–886 (2002). doi: 10.1006/jhev.2002.0605; pmid: 12473487
23. E. Yong, "A scientist witnessed poachers killing a chimp" *The Atlantic*, 1 August 2019; https://www.theatlantic.com/science/archive/2019/08/death-chimpanzee/595303/.
24. K. B. Potts, C. A. Chapman, J. S. Lwanga, Floristic heterogeneity between forested sites in Kibale National Park, Uganda: Insights into the fine-scale determinants of density in a large-bodied frugivorous primate. *J. Anim. Ecol.* **78**, 1269–1277 (2009). doi: 10.1111/j.1365-2656.2009.01578.x; pmid: 19523110
25. K. B. Potts, D. P. Watts, R. W. Wrangham, Comparative feeding ecology of two communities of chimpanzees (*Pan troglodytes*) in Kibale National Park, Uganda. *Int. J. Primatol.* **32**, 669–690 (2011). doi: 10.1007/s10764-011-9494-y
26. D. P. Watts, K. B. Potts, J. S. Lwanga, J. C. Mitani, Diet of chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda, 1. Diet composition and diversity. *Am. J. Primatol.* **74**, 114–129 (2012). doi: 10.1002/ajp.21016; pmid: 22109938
27. J. C. Mitani, D. P. Watts, S. J. Amsler, Lethal intergroup aggression leads to territorial expansion in wild chimpanzees. *Curr. Biol.* **20**, R507–R508 (2010). doi: 10.1016/j.cub.2010.04.021; pmid: 20620900
28. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford Univ. Press, 2018).
29. D. A. Levitis, O. Burger, L. B. Lackey, The human post-fertile lifespan in comparative evolutionary context. *Evol. Anthropol.* **22**, 66–79 (2013). doi: 10.1002/evan.21332; pmid: 23585379
30. J. F. O'connell, K. Hawkes, N. G. Blurton Jones, Grandmothering and the evolution of *Homo erectus*. *J. Hum. Evol.* **36**, 461–485 (1999). doi: 10.1006/jhev.1998.0285; pmid: 10222165
31. C. Boesch, H. Boesch-Achermann, *The Chimpanzees of the Taï Forest: Behavioural Ecology and Evolution* (Oxford Univ. Press, 2000).
32. R. E. Moreau, The distribution of chimpanzees in Tanganyika Territory. *Tanganyika Notes Rec.* **14**, 52–55 (1942).
33. T. T. Struhsaker, *Ecology of an African Rain Forest: Logging in Kibale and the Conflict Between Conservation and Exploitation* (Univ. Press of Florida, 1997).
34. V. Reynolds, *The Chimpanzees of the Budongo Forest: Ecology, Behaviour and Conservation* (Oxford Univ. Press, 2005).
35. T. Matsuzawa, T. Humle, Y. Sugiyama, Eds., *The Chimpanzees of Bossou and Nimba*, Primatology Monographs (Springer, 2011).
36. M. Emery Thompson *et al*., Risk factors for respiratory illness in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). *R. Soc. Open Sci.* **5**, 180840 (2018). doi: 10.1098/rsos.180840; pmid: 30839693
37. J. D. Negrey *et al*., Simultaneous outbreaks of respiratory disease in wild chimpanzees caused by distinct viruses of human origin. *Emerg. Microbes Infect.* **8**, 139–149 (2019). doi: 10.1080/22221751.2018.1563456; pmid: 30866768
38. J. D. Negrey *et al*., Viruses associated with ill health in wild chimpanzees. *Am. J. Primatol.* **84**, e23358 (2022). doi: 10.1002/ajp.23358; pmid: 35015311
39. M. N. Muller, R. W. Wrangham, Mortality rates among Kanyawara chimpanzees. *J. Hum. Evol.* **66**, 107–114 (2014). doi: 10.1016/j.jhevol.2013.10.004; pmid: 24374229
40. A. M. Bronikowski *et al*., Female and male life tables for seven wild primate species. *Sci. Data* **3**, 160006 (2016). doi: 10.1038/sdata.2016.6; pmid: 26928014
41. P. B. Medawar, *An Unsolved Problem of Biology* (H. K. Lewis and Co., ed. 1, 1952).
42. P. S. Kim, J. S. McQueen, J. E. Coxworth, K. Hawkes, Grandmothering drives the evolution of longevity in a probabilistic model. *J. Theor. Biol.* **353**, 84–94 (2014). doi: 10.1016/j.jtbi.2014.03.011; pmid: 24637003
43. A. R. Rogers, Why menopause? *Evol. Ecol.* **7**, 406–420 (1993). doi: 10.1007/BF01237872
44. K. Hill, A. M. Hurtado, The evolution of premature reproductive senescence and menopause in human females: An evaluation of the "grandmother hypothesis". *Hum. Nat.* **2**, 313–350 (1991). doi: 10.1007/BF02692196; pmid: 24222339
45. T. Matsumoto, S. Hanamura, T. Kooriyama, T. Hayakawa, E. Inoue, Female chimpanzees giving first birth in their natal group in Mahale: Attention to incest between brothers and sisters. *Primates* **62**, 279–287 (2021). doi: 10.1007/s10329-020-00886-3; pmid: 33442833
46. A. C. Pisor, C. T. Ross, Distinguishing intergroup and long-distance relationships. *Hum. Nat.* **33**, 280–303 (2022). doi: 10.1007/s12110-022-09431-1; pmid: 36181615
47. K. R. Hill, B. M. Wood, J. Baggio, A. M. Hurtado, R. T. Boyd, Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PLOS ONE* **9**, e102806 (2014). doi: 10.1371/journal.pone.0102806; pmid: 25047714
48. K. E. Langergraber, J. C. Mitani, L. Vigilant, The limited impact of kinship on cooperation in wild chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7786–7790 (2007). doi: 10.1073/pnas.0611449104; pmid: 17456600
49. A. A. Sandel, K. E. Langergraber, J. C. Mitani, Adolescent male chimpanzees (*Pan troglodytes*) form social bonds with their brothers and others during the transition to adulthood. *Am. J. Primatol.* **82**, e23091 (2020). doi: 10.1002/ajp.23091; pmid: 31903634
50. C. M. Murray, M. A. Stanton, E. V. Lonsdorf, E. E. Wroblewski, A. E. Pusey, Chimpanzee fathers bias their behaviour towards their offspring. *R. Soc. Open Sci.* **3**, 160441 (2016). doi: 10.1098/rsos.160441; pmid: 28018626
51. C. Boesch, J. Lehmann, G. Fickenscher, Kin biased investment in wild chimpanzees. *Behaviour* **143**, 931–955 (2006). doi: 10.1163/156853906778623635
52. M. A. Cant, R. A. Johnstone, Reproductive conflict and the separation of reproductive generations in humans. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5332–5336 (2008). doi: 10.1073/pnas.0711911105; pmid: 18378891
53. S. R. Copeland *et al*., Strontium isotope evidence for landscape use by early hominins. *Nature* **474**, 76–78 (2011). doi: 10.1038/nature10149; pmid: 21637256
54. L. Skov *et al*., Genetic insights into the social organization of Neanderthals. *Nature* **610**, 519–525 (2022). doi: 10.1038/s41586-022-05283-y; pmid: 36261548
55. H. P. Alvarez, "Residence groups among hunter-gatherers: a view of the claims and evidence for patrilocal bands" in *Kinship and Behavior in Primates*, B. Chapais, C. M. Berman, Eds. (Oxford Univ. Press, 2004), pp. 420–442.
56. S. Ellis *et al*., Patterns and consequences of age-linked change in local relatedness in animal societies. *Nat. Ecol. Evol.* **6**, 1766–1776 (2022). doi: 10.1038/s41559-022-01872-2; pmid: 36163259
57. D. P. Croft *et al*., Reproductive conflict and the evolution of menopause in killer whales. *Curr. Biol.* **27**, 298–304 (2017). doi: 10.1016/j.cub.2016.12.015; pmid: 28089514
58. R. Mace, A. Alvergne, Female reproductive competition within families in rural Gambia. *Proc. Biol. Sci.* **279**, 2219–2227 (2012). doi: 10.1098/rspb.2011.2424; pmid: 22258635
59. H. J. Nichols, K. Arbuckle, K. Fullard, W. Amos, Why don't long-finned pilot whales have a widespread postreproductive lifespan? Insights from genetic data. *Behav. Ecol.* **31**, 508–518 (2020). doi: 10.1093/beheco/arz211
60. A. M. John, *The Plantation Slaves of Trinidad, 1783–1816: A Mathematical and Demographic Enquiry* (Cambridge Univ. Press, 1988).
61. J. S. Lwanga, Localized tree mortality following the drought of 1999 at Ngogo, Kibale National Park, Uganda. *Afr. J. Ecol.* **41**, 194–196 (2003). doi: 10.1046/j.1365-2028.2003.00428.x
62. K. B. Potts, D. P. Watts, K. E. Langergraber, J. C. Mitani, Long-term trends in fruit production in a tropical forest at Ngogo, Kibale National Park, Uganda. *Biotropica* **52**, 521–532 (2020). doi: 10.1111/btp.12764; pmid: 33692573
63. J. S. Lwanga, Forest succession in Kibale National Park, Uganda: Implications for forest restoration and management. *Afr. J. Ecol.* **41**, 9–22 (2003). doi: 10.1046/j.1365-2028.2003.00400.x
64. J. S. Lwanga, T. M. Butynski, T. T. Struhsaker, Tree population dynamics in Kibale National Park, Uganda 1975–1998. *Afr. J. Ecol.* **38**, 238–247 (2000). doi: 10.1046/j.1365-2028.2000.00244.x
65. B. M. Wood, D. P. Watts, J. C. Mitani, K. E. Langergraber, Favorable ecological circumstances promote life expectancy in chimpanzees similar to that of human hunter-gatherers.

J. Hum. Evol. **105**, 41–56 (2017). doi: 10.1016/j.jhevol.2017.01.003; pmid: 28366199
66. K. Hill *et al*., Mortality rates among wild chimpanzees. *J. Hum. Evol.* **40**, 437–450 (2001). doi: 10.1006/jhev.2001.0469; pmid: 11322804
67. M. N. Muller, S. F. Lipson, Diurnal patterns of urinary steroid excretion in wild chimpanzees. *Am. J. Primatol.* **60**, 161–166 (2003). doi: 10.1002/ajp.10103; pmid: 12910467
68. J. L. Brown, S. B. Citino, M. Bush, J. Lehnhardt, L. G. Phillips, Cyclic patterns of luteinizing hormone, follicle-stimulating hormone, inhibin, and progesterone secretion in the Asian elephant (*Elephas maximus*). *J. Zoo Wildl. Med.* **22**, 49–57 (1991).
69. B. Hauser, T. Deschner, C. Boesch, Development of a liquid chromatography-tandem mass spectrometry method for the determination of 23 endogenous steroids in small quantities of primate urine. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **862**, 100–112 (2008). doi: 10.1016/j.jchromb.2007.11.009; pmid: 18054529
70. R. C. Miller *et al*., Comparison of specific gravity and creatinine for normalizing urinary reproductive hormone concentrations. *Clin. Chem.* **50**, 924–932 (2004). doi: 10.1373/clinchem.2004.032292; pmid: 15105350
71. K. K. Walker, C. S. Walker, J. Goodall, A. E. Pusey, Maturation is prolonged and variable in female chimpanzees. *J. Hum. Evol.* **114**, 131–140 (2018). doi: 10.1016/j.jhevol.2017.10.010; pmid: 29447755
72. L. J. Peacock, C. M. Rogers, Gestation period and twinning in chimpanzees. *Science* **129**, 959 (1959). doi: 10.1126/science.129.3354.959.a; pmid: 13646623
73. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. arXiv:1406.5823 [stat.CO] (2014).
74. S. G. Luke, Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* **49**, 1494–1502 (2017). doi: 10.3758/s13428-016-0809-y; pmid: 27620283
75. A. Kuznetsova, P. B. Brockhoff, R. H. Christensen, lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017). doi: 10.18637/jss.v082.i13
76. R. Miller, F. Plessow, Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations. *Psychoneuroendocrinology* **38**, 941–946 (2013). doi: 10.1016/j.psyneuen.2012.09.013; pmid: 23063878
77. G. E. Box, D. R. Cox, An analysis of transformations. *J. R. Stat. Soc. B* **26**, 211–243 (1964). doi: 10.1111/j.2517-6161.1964.tb00553.x
78. M. B. Wilk, R. Gnanadesikan, Probability plotting methods for the analysis of data. *Biometrika* **55**, 1–17 (1968). doi: 10.2307/2334448; pmid: 5661047
79. M. Kutner, C. Nachtsheim, J. Neter, *Applied Linear Regression Models* (McGraw-Hill/Irwin, ed. 4, 2004).
80. S. N. Wood, *Generalized Additive Models: An Introduction with R* (CRC Press, 2017).
81. A. Rohatgi, WebPlotDigitizer: Version 4.6 (2017); https://automeris.io/WebPlotDigitizer/.
82. W. Hobson, F. Coulston, C. Faiman, J. S. Winter, F. Reyes, Reproductive endocrinology of female chimpanzees: A suitable model of humans. *J. Toxicol. Environ. Health* **1**, 657–668 (1976). doi: 10.1080/15287397609529364; pmid: 1263283
83. J. Goodall, *The Chimpanzees of Gombe: Patterns of Behavior* (Belknap Press, 1986).
84. N. Blurton Jones, *Demography and Evolutionary Ecology of Hadza Hunter-Gatherers*, vol. 71 of *Cambridge Studies in Biological and Evolutionary Anthropology* (Cambridge Univ. Press, 2016).
85. M. Gurven, H. Kaplan, Longevity among hunter-gatherers: A cross-cultural examination. *Popul. Dev. Rev.* **33**, 321–365 (2007). doi: 10.1111/j.1728-4457.2007.00171.x
86. M. Cords, S. Chowdhury, Life history of *Cercopithecus mitis stuhlmanni* in the Kakamega Forest, Kenya. *Int. J. Primatol.* **31**, 433–455 (2010). doi: 10.1007/s10764-010-9405-7
87. A. M. Bronikowski *et al*., Aging in the natural world: Comparative data reveal similar mortality patterns across primates. *Science* **331**, 1325–1328 (2011). doi: 10.1126/science.1201571; pmid: 21393544
88. Y. Takahata *et al*., Reproduction of wild Japanese macaque females of Yakushima and Kinkazan Islands: A preliminary report. *Primates* **39**, 339–349 (1998). doi: 10.1007/BF02573082
89. C. Packer, M. Tatar, A. Collins, Reproductive cessation in female mammals. *Nature* **392**, 807–811 (1998). doi: 10.1038/33910; pmid: 9572138
90. S. Ichino *et al*., Lifespan and reproductive senescence in a free-ranging ring-tailed lemur (*Lemur catta*) population at Berenty, Madagascar. *Folia Primatol. (Basel)* **86**, 134–139 (2015). doi: 10.1159/000368670; pmid: 26022309
91. M. S. MacDonald Pavelka, L. M. Fedigan, S. Zohar, Availability and adaptive value of reproductive and postreproductive Japanese macaque mothers and grandmothers. *Anim. Behav.* **64**, 407–414 (2002). doi: 10.1006/anbe.2002.3085
92. N. Koyama, Y. Takahata, M. A. Huffman, K. Norikoshi, H. Suzuki, Reproductive parameters of female Japanese macaques: Thirty years data from the Arashiyama troops, Japan. *Primates* **33**, 33–47 (1992). doi: 10.1007/BF02382761
93. N. Howell, *Demography of the Dobe! Kung* (Routledge, ed. 2, 2017).
94. K. Hill, A. M. Hurtado, *Ache Life History: The Ecology and Demography of a Foraging People* (Aldine de Gruyter, 1996).
95. E. A. G. von Hofsten, H. Lundström, *Swedish Population History: Main Trends from 1750 to 1970* (Stockholm: Statistiska centralbyr an:[LiberFörlag], 1976).

ACKNOWLEDGMENTS

We are grateful to the Makerere University Biological Field Station, Uganda Wildlife Authority, and Uganda National Council for Science and Technology for sponsoring our research in Uganda. We thank the Ngogo students, postdocs, and other researchers who have contributed to the chimpanzee demographic database and the late J. Lwanga and S. Angedakin, who served as camp managers for the Ngogo Chimpanzee Project. We are indebted to R. Davenport for sample collection, and to project field assistants C. Aligarya, C. Birungi, D. Kalunga, B. Kamugyisha, D. Kamweri, A. Magoba, G. Mbabazi, L. Ngandizi, A. Tumusiime, and A. Twineomujuni. **Funding:** We are grateful to the funders of research at Ngogo, including the US National Science Foundation (BCS-9253590, IOS-0516644, BCS-0215622, BCS-9253590, BCS-1613393), the National Institute on Aging and the Office for Research on Women's Health (R01-AG049395), the Fulbright Foundation, the Explorer's Club, the National Geographic Society (9824-15), the Detroit Zoological Society, Underdog Films, Silverback Films, University of Michigan, Yale University, Arizona State University President's Strategic Initiative Fund, the Institute of Human Origins, the Max Planck Society, the Wenner-Gren Foundation, and the L.S.B. Leakey Foundation. **Author contributions:** Conceptualization: J.C.M., D.P.W., K.E.L., and B.M.W. Field data collection: J.C.M., D.P.W., K.E.L., S.G., and J.D.N. Hormone measurement: J.L.B., T.D., M.E.T., S.G., and J.D.N. Demographic data analyses: B.M.W. Categorical hormone data analyses: J.D.N. and B.M.W. Continuous hormone data analyses: B.M.W. Writing of original draft: B.M.W. and K.E.L. Major contributions to subsequent drafts: B.M.W., K.E.L., M.E.T., J.C.M., D.P.W., and T.D. Editing: All authors. **Competing interests:** The authors have no competing interests to declare. **Data and materials availability:** The data and code ("all_functions.R") necessary to reproduce all the analyses in this paper are provided in the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add5473
Figs. S1 to S5
Tables S1 to S7
MDAR Reproducibility Checklist
Data S1 to S16

Submitted 17 June 2022; accepted 5 September 2023
10.1126/science.add5473

RESEARCH ARTICLE

NEUROSCIENCE

Cross-modal representation of identity in the primate hippocampus

Timothy J. Tyree[1,2], Michael Metke[1,3], Cory T. Miller[1,3]*

Faces and voices are the dominant social signals used to recognize individuals among primates. Yet, it is not known how these signals are integrated into a cross-modal representation of individual identity in the primate brain. We discovered that, although single neurons in the marmoset hippocampus exhibited selective responses when presented with the face or voice of a specific individual, a parallel mechanism for representing the cross-modal identities for multiple individuals was evident within single neurons and at the population level. Manifold projections likewise showed the separability of individuals as well as clustering for others' families, which suggests that multiple learned social categories are encoded as related dimensions of identity in the hippocampus. Neural representations of identity in the hippocampus are thus both modality independent and reflect the primate social network.

Navigating complex primate societies relies on learning the identity of each individual in the group and their respective social relationships (*1*). Neurons in the brains of primates and other mammals selectively respond to the identity when viewing the face or hearing the voice of a specific individual as unimodal signals (*2*–*8*). However, data showing that single neurons are responsive to both the face and voice of an individual—a cross-modal representation of identity—are limited to "concept cells" in the human hippocampus (*9*–*11*). These neurons are notable for several reasons, including their putative role in memory functions (*12*) and potential uniqueness to humans (*13*). We investigated whether cross-modal representations of identity are evident in the hippocampus of marmoset monkeys by recording single hippocampal neurons (*14*) while presenting subjects with multiple exemplars of individual marmoset faces (from different viewpoints) and voices as unimodal stimuli (*4*), as well as concurrently by presenting the faces and voices from the same or different individuals—i.e., match versus mismatch (MvMM). Visual stimuli were presented from a monitor directly in front of the animal while a speaker positioned directly below the screen broadcast the acoustic stimuli. Subjects were only presented with familiar conspecifics housed in the same colony who differed in their respective social relatedness (*11*).

Identity-selective neurons

To first test whether cross-modal representations of identity are evident in the hippocampus of a nonhuman primate, we performed the same receiver operator characteristic (ROC) selectivity analysis described previously in humans (*9*–*11*) and detected a population of cross-modal invariant neurons for individual identity when observing marmoset faces or voices (Fig. 1A and fig. S1A), as well as neurons selective for individual identity when viewing only their faces (Fig. 1B and fig. S1B) or hearing only their voices (Fig. 1C and fig. S1C). These identity neurons were confirmed in all hippocampal subfields (Fig. 1D). Notably, only neurons that exhibited a mean peak firing rate 2 SDs above baseline qualified for this analysis, which supports $P < 0.01$, and differed from the 5 SDs used previously in humans (*9*–*11*). Responses were determined from the mean firing rate during a 500-ms continuous sliding window maximized over the duration of the 3500-ms stimulus. Overall, we observed that $N = 148$ (9.2%) of $N = 1602$ qualifying neurons demonstrated selectivity for a single preferred individual (Fig. 1E), with different neurons selective for faces ($N = 52$), voices ($N = 39$), or both faces and voices ($N = 57$) (Fig. 1F). The mean area under the ROC curve (AUC) of identity neurons (AUC = 0.902 ± 0.014) was significantly above chance ($P < 0.001$) (Fig. 1G). Although these neurons in marmosets were overall less selective than in humans (*9*–*11*), this disparity may reflect species differences in hippocampus properties that affect neural coding mechanisms for identity. Baseline hippocampal activity, for example, was considerably higher in the current study (mean 6.47 Hz; $N = 2358$ neurons) (fig. S2) than has been reported in humans (*15*), although a more comprehensive comparative analysis of physiological differences is needed to better understand how such differences affect hippocampal functions.

Analysis of eye movements (Fig. 1H) revealed that marmosets' visual behavior and neural activity were differentially affected by modality and identity. Marmosets exhibited significantly shorter fixations ($P < 0.001$, $N_{fixations} = 18{,}965$) (Fig. 1I) and significantly more saccades ($P < 0.001$, $N_{saccades} = 2203$) during trials with face-only relative to the voice-only trials (Fig. 1J). These monkeys were also highly focused on faces during stimulus presentations, with faces accounting for 77.9% of viewing time and eyes specifically accounting for 37.6% of viewing time. The firing rate of identity neurons was significantly greater than the remaining neurons when subjects were looking at the eyes or face (both $P < 0.001$) (Fig. 1K). This was not, however, a broad attentional effect (*16*) because the firing rate of simultaneously recorded nonidentity neurons did not show the same increased firing rate when gazing at faces or eyes.

Multiple identities are represented in single neurons

A potential parallel mechanism to highly selective concept cells is for individual cells to contribute to multiple functions (*17*, *18*), such as single neurons being sensitive to the cross-modal identity of multiple conspecifics. Hippocampal neurons are sensitive to mismatches between the features of a particular stimulus and a previously learned category (*19*, *20*). To test whether a similar mechanism is evident for the learned social identities of conspecifics in marmoset hippocampus, we tested whether neurons would respond differently when simultaneously observing the face and voice from the same (identity match) or different (identity mismatch) individuals. By presenting a face and voice in all identity MvMM trials, we controlled for the potential effects of multimodal integration (fig. S3A) and instead tested whether a subordinate category, identity, elicited changes in neural activity. Indeed, a subpopulation of units, MvMM neurons, exhibited a significant firing rate preference for either match trials (Fig. 2A) or mismatch trials (Fig. 2B), with some neurons modulated only by this category distinction (Fig. 2A) and others more generally stimulus driven (Fig. 2B). Overall, 21.7% of neurons ($N = 511$ of 2358) exhibited a significant response during MvMM trials, with significantly more units exhibiting a higher firing rate during match ($N = 401$) than mismatch ($N = 110$) trials ($P < 0.001$) (Fig. 2C and fig. S3B). MvMM neurons were largely distinct from the identity neurons described above (Fig. 2D and fig. S3C). Notably, 56% of the neurons observed in both populations whose anatomical location could be confirmed were recorded in CA1. In contrast to identity neurons, MvMM neurons were biased to CA1 (Fig. 2E), with $N = 155$ (44.3%) out of 350 neurons confirmed in the CA1 qualifying as MvMM neurons. In CA1, significantly more MvMM neurons ($N = 129/155$, 83.2%) preferred match

[1]Cortical Systems and Behavior Laboratory, University of California San Diego, La Jolla, CA 92039, USA. [2]Department of Physics, University of California San Diego, La Jolla, CA 92039, USA. [3]Neurosciences Graduate Program, University of California San Diego, La Jolla, CA 92039, USA.
*Corresponding author. Email: corymiller@ucsd.edu

trials to mismatch trials ($P < 0.001$). MvMM neurons exhibited significantly higher median firing rate while the subject was looking at the eyes or face ($P < 0.001$, $N = 511$) (Fig. 2F). Marmosets exhibited significantly more saccadic eye movements during mismatch trials (Fig. 2G), and this difference in behavior was most prominent 1 to 2 s after stimulus onset ($P < 0.05$, $N_{saccades} = 4603$) (Fig. 2H).

Encoding cross-modal identity in neuron populations

These findings suggest that two seemingly distinct mechanisms for representing cross-modal identity are evident in primate hippocampus. We conjectured that more temporally selective coding mechanisms in hippocampus may inform how these two processes for encoding identity are integrated at a population level. Therefore, we developed an algorithm to identify intervals of time during which individual

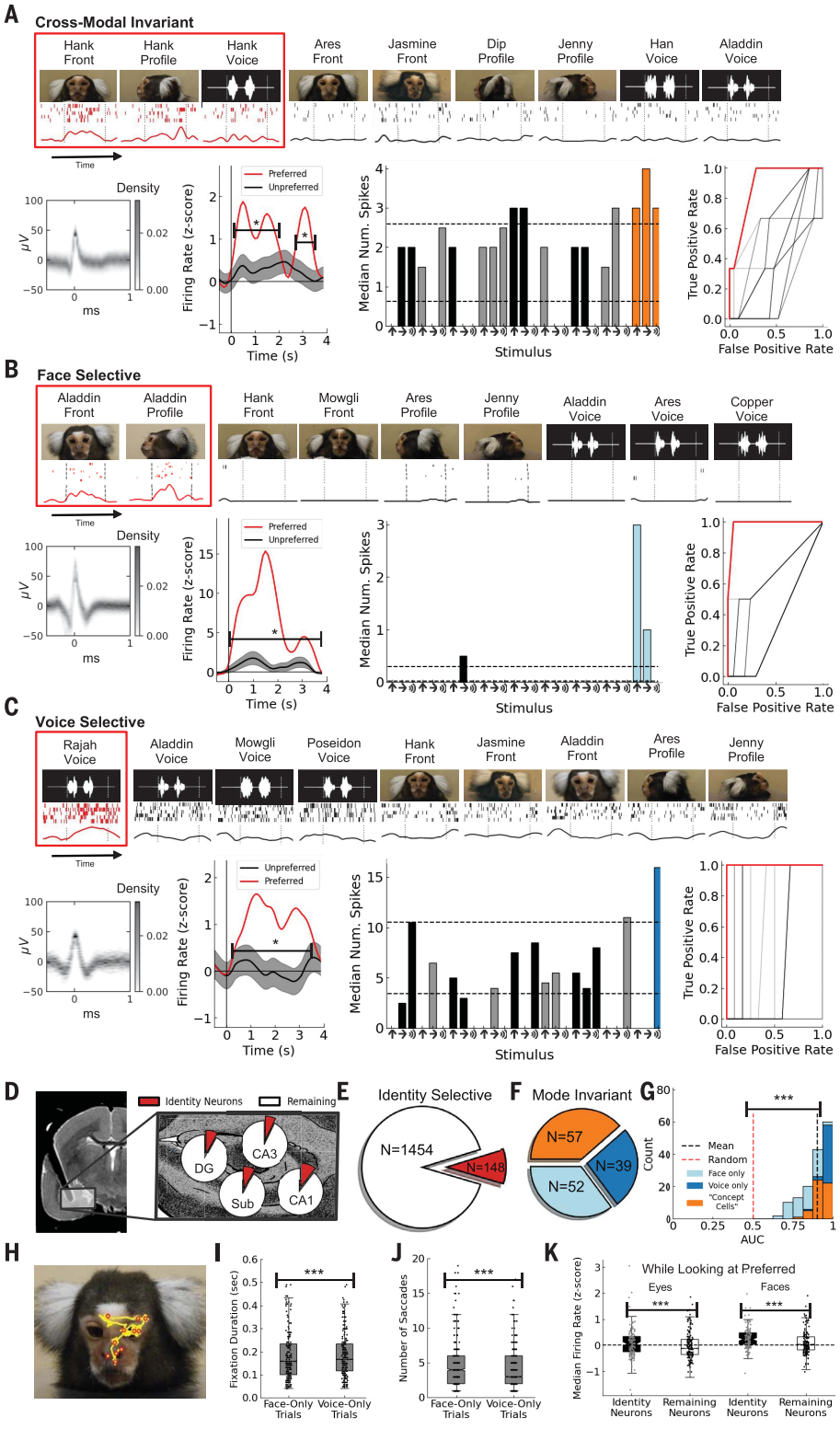


Fig. 1. Putative concept cells in marmoset hippocampus. (**A** to **C**) Top row: Subset of stimuli shown above raster and peristimulus time histogram (PSTH). Bottom row: Spike waveform density; normalized PSTH to all stimuli (preferred: red, nonpreferred: black), indicated are time points that show significant difference ($P < 0.05$); median number of spikes for unimodal stimuli (gray/black indicate nonpreferred individuals); ROC curve (shuffled controls shown in black). Exemplar identity neurons responding selectively to the face and voice of a preferred conspecific (red) (A), the face only (B), and the voice only (C) are shown. (**D**) Anatomical distribution of identity neurons (red) in hippocampal subfields relative to neurons remaining that responded to any stimulus (white). Black shadow indicates the electrode array track with magnetic resonance imaging distortion artifact. DG, dentate gyrus; Sub, subiculum. (**E**) Pie chart showing the abundance of identity neurons in red with the number of remaining neurons that qualified for the ROC selectivity analysis in white. (**F**) Mode distribution of identity neurons. Modes included face (light blue), voice (dark blue), and both (orange). (**G**) Histogram showing the distribution of AUCs comparable with red ROC curves in (A) to (C). Colors are as in (F). Black dotted line is the mean, and red dotted line is the mean of 10,000 random shuffles of the labels. (**H**) Exemplar eye movements (yellow) with fixations indicated (red). (**I**) Distribution of eye fixation durations for unimodal trials. (**J**) Distribution of apparent saccade number for unimodal trials. (**K**) Distribution of median firing rates while observer was looking at eyes (left) and face (right) for identity neurons (black) versus remaining neurons (white). Significant median differences, ***$P < 0.001$.

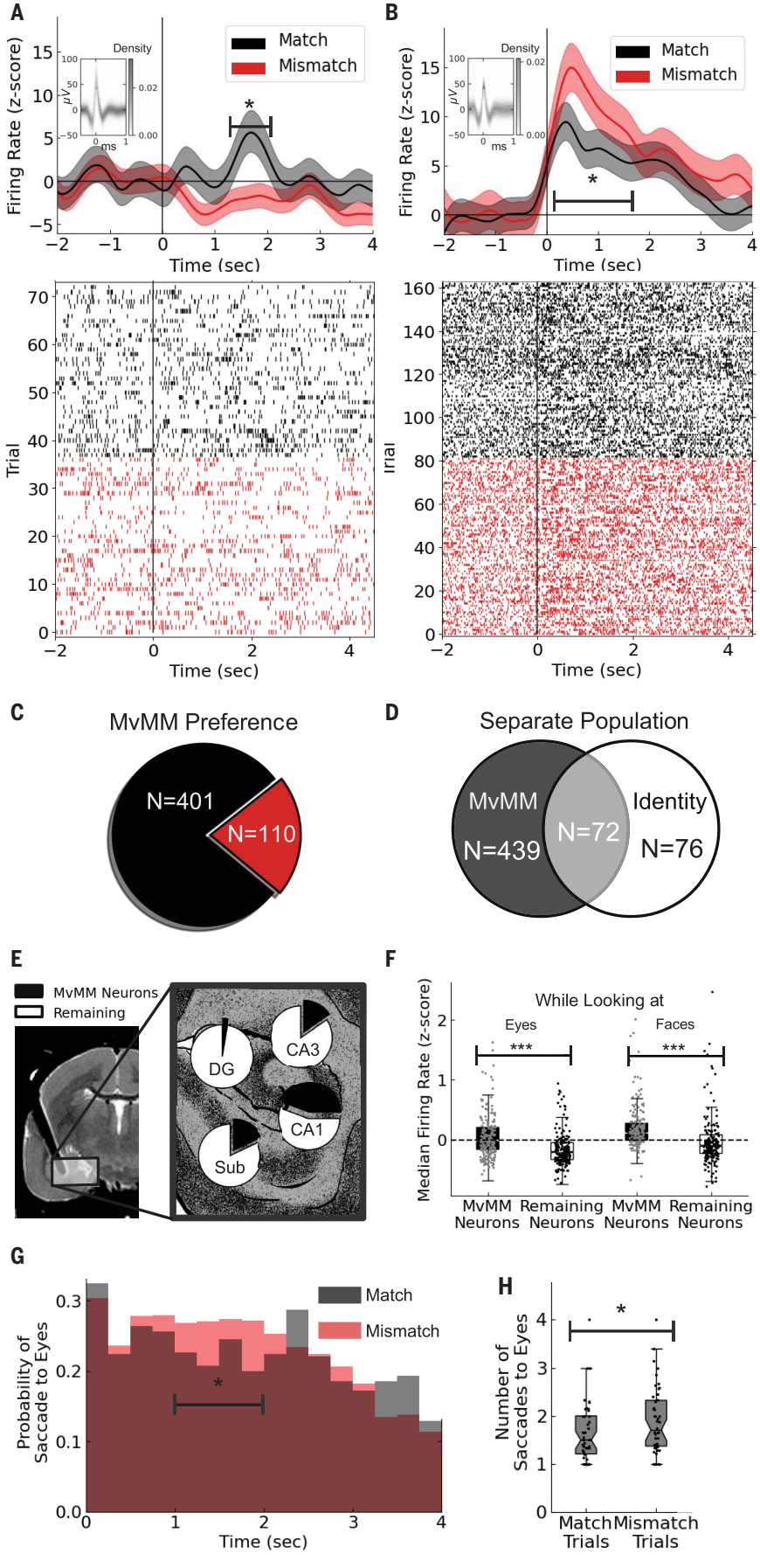


Fig. 2. Single neurons in hippocampus represent multiple individuals. (**A** and **B**) PSTH normalized by the prestimulus baseline (top) and spike raster (bottom) for two exemplar MvMM neurons. Black indicates match trials, and red indicates mismatch trials. Vertical line indicates stimulus onset. Inset shows spike waveform density. Significant time points, *P < 0.05. Exemplar neuron with higher firing rate for match (A) and mismatch (B) trials. (**C**) Pie chart showing the number of neurons that responded significantly more for match (black) or mismatch (red) trials. (**D**) Venn diagram showing the number of MvMM neurons (black) in common with identity neurons (red). (**E**) Relative abundance of MvMM neurons in each hippocampal subfield. (**F**) Distribution of median firing rate while looking at the eyes (left) and face (right) for MvMM neurons (black) versus remaining neurons (white). (**G**) Probability density of saccadic eye movements directed toward the eyes for match (black) and mismatch (red) trials. Indicated are the time points in (H). (**H**) Distribution of apparent number of saccades to eyes. Significant median differences, *P < 0.05.

neurons exhibited significant differences in median firing rate for a specific category ($P < 0.05$), which we labeled as predictive time bins (fig. S4). This algorithm was applied to all neurons in the population, not only those classified as identity-selective or MvMM neurons. We first identified predictive time bins selective for specific individuals when observing their face or voice. A pair of exemplar neurons that exhibited separate predictive time bins for two different individuals is shown (Fig. 3A and fig. S5). Out of 2358 hippocampal neurons, 1634 (69.3%) exhibited at least one identity-specific predictive time bin, with most exhibiting predictive time bins for two or more individuals (Fig. 3B). Identity-specific predictive time bins exhibited a mean AUC (0.802 ± 0.003) that was significantly above chance ($P < 0.001$, $N_{bins} = 3958$) (Fig. 3C). Neurons that had identity-specific predictive time bins exhibited a significantly greater median firing rate when subjects were looking at the face of a preferred individual ($P < 0.001$) (fig. S6A), although significant suppression was observed relative to

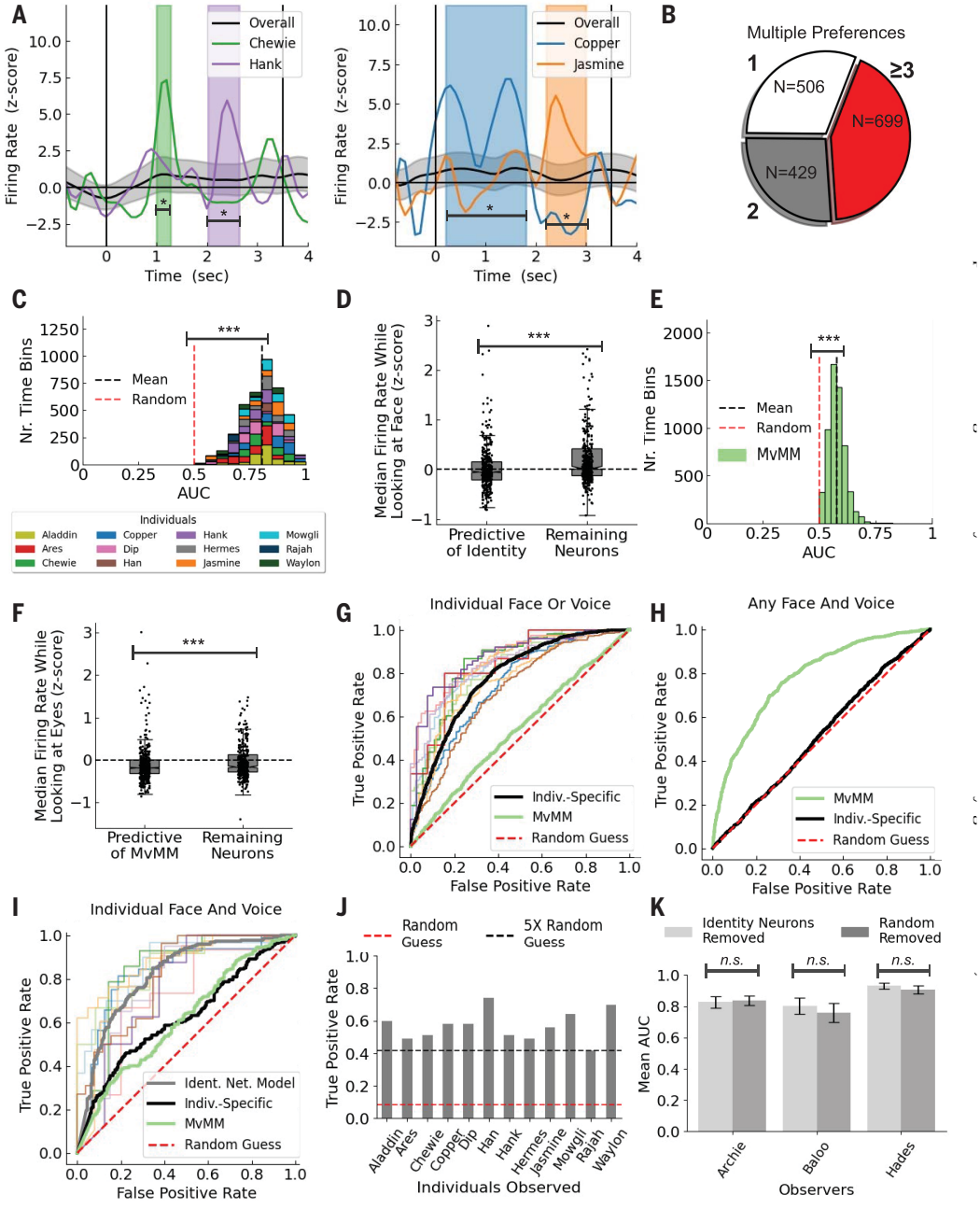


Fig. 3. Cross-modal decoding of identity. (**A**) PSTH of two exemplar predictive neurons. Colored traces average over trials involving preferred individual, and the gray shaded regions indicate 95% confidence intervals. Colored regions indicate identity-specific time bins. (**B**) Pie chart showing number of identity-specific predictive neurons that prefer one (white), two (gray), and three or more individuals (red). (**C**) Histogram showing AUC distribution of identity-specific time bins with colors indicating preferred individuals in legend. Dotted lines indicate the mean (black) and the control (red). (**D**) Distribution of median firing rates while the observer was looking at the face for the identity-specific predictive neurons compared with the remaining neurons. (**E**) Histogram showing AUC distribution of MvMM time bins. Dashed lines indicate the mean (black) and the control (red). (**F**) Distribution of median firing rates while the observer was looking at the face for the MvMM predictive neurons compared with the remaining neurons. (**G**) ROC curves for the detection of face or voice of individuals. Firing rates were considered from MvMM time bins (green, AUC = 0.536) and identity-specific time bins (black, AUC = 0.779) similarly averaged over individuals. Thinner colored lines indicate individuals as in (C). (**H**) ROC curves for the detection of match trials. Firing rates from MvMM time bins (green, AUC = 0.782) and from identity-specific time bins (black, AUC = 0.516). (**I**) ROC curves for the detection of both face and voice of individuals from same 19 recording sessions as in (G) and (H). Firing rates from MvMM time bins (green, AUC = 0.615), identity-specific time bins (black, AUC = 0.622), and the INM (gray, AUC = 0.818) are similarly averaged over individuals. Results of the INM for individuals are shown by thin lines colored as in the legend of (C). Red dotted line indicates random as in (G) and (H). (**J**) Bar plot showing true positive rates predicted by a winner-take-all model that considered predictions from the INM specific to 12 individuals. Indicated is the mean of the shuffled labels (red) and 5× that value (black). Bar plots summarize the trials from the testing sets of 33 recording sessions ($N_{trials} = 454$). (**K**) Bar plot showing mean AUC with identity neurons removed (light gray) versus the control randomly removing an equal number of bins from the remaining cells (dark gray). Uncertainty indicates 95% confidence of the mean. No significant difference was observed across recording sessions for any of the three qualifying subjects (Archie, $P = 0.81$, $N_{identities} = 14$; Baloo, $P = 0.58$, $N_{identities} = 9$; Hades, $P = 0.50$, $N_{identities} = 12$). ***$P < 0.001$; n.s., not significant.

the other neurons when normalizing by the background, which was averaged from t = −0.8 s to t = −0.3 s (P < 0.001) (Fig. 3D). We applied the same algorithm to test for predictive time bins that distinguished MvMM trials and found a similar result (Fig. 3E), with 1455 neurons exhibiting MvMM predictive time bins. Neurons with predictive time bins for MvMM exhibited a significantly greater median firing rate when subjects looked at the face (P < 0.001) (fig. S6B), although significant suppression was observed relative to the

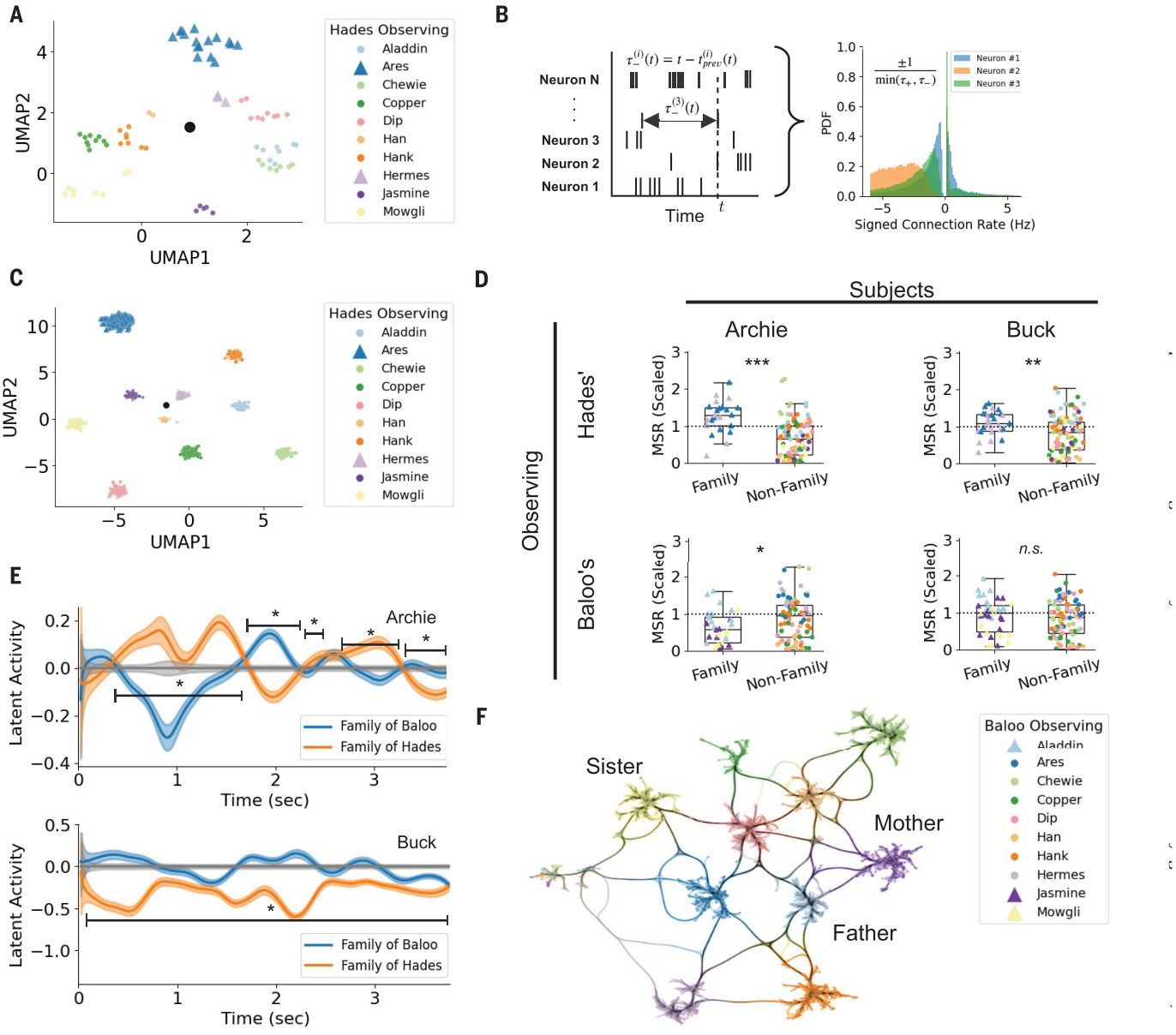


Fig. 4. Cross-modal representation of identity using rate and event codes. (**A**) Two-dimensional manifold projection of our rate-coded representation computed from firing rates of identity-specific time bins. One identity-match trial was equivalent to one presentation of the stimulus as face and voice matched. Each identity-match trial represented a different, randomly selected face and/or voice stimulus. Firing rates were computed from each identity-specific predictive time bins and then concatenated into a feature vector for each identity-match trial. This feature vector was then projected onto the manifold and plotted as one symbol per identity-match trial in one of the scatter plots. Indicated is the mean (black). Colors in legend correspond to individuals. UMAP, uniform manifold approximation and projection. (**B**) Schematic illustrating the hindsight delay to a given neuron (left), used to generate histograms of signed connection rates to three neurons (right). PDF, probability density function. (**C**) Two-dimensional manifold projection of our event-coded representation of identity computed as the manifold projection of signed connection rates of all neurons in the same exemplar recording session. One symbol represents one spike. Indicated is the mean (black). (**D**) Boxplots of MSR showing significantly different values when subjects observed family of other subjects. Shown is Archie observing family of Hades (top left, P < 0.001, $N_{identities}$ ≥ 23), Buck observing family of Hades (top right, P = 0.003, $N_{identities}$ ≥ 26), Archie observing family of Baloo (bottom left, P = 0.017, $N_{identities}$ ≥ 30), and Buck observing family of Baloo (bottom right, P = 0.828, $N_{identities}$ ≥ 37). Significance was computed according to Student's t test. (**E**) Latent activity averaged over all recording sessions from subjects Archie (left) and Buck (right). Colors indicate average over the family of Baloo (blue) and Hades (orange) relative to all conspecifics (gray). Shaded regions indicate 95% confidence of the mean estimated through bootstrap. (**F**) Graph of connections bundled between individuals. Triangles in legend indicate family members as in (A) and (C).

other neurons when normalizing by the same background ($P < 0.001$) (Fig. 3F). Instances of face and eye viewing were highly variable and not limited to the timing of predictive time bins, which suggests that attentional effects from visual behavior were not likely driving neural activity during these periods (fig. S7). We observed considerable overlap between neurons with identity-specific and MvMM predictive time bins because 82.2% ($N = 1196$; fig. S8A) exhibited predictive time bins in both analyses.

Identity network model

We developed a stable neural decoder by combining the firing rates of predictive time bins using an ensemble of gradient-boosted decisions trees (*21*). When using identity-specific time bins, we could reliably decode the identity of all marmosets when subjects observed their face or voice (accuracy: 77.4%) (Fig. 3G). The same approach could successfully decode MvMM trials when using MvMM time bins (accuracy: 75.7%) (Fig. 3H). The two kinds of decoders used mostly different time points, with only 24.6% ± 1.5% of identity-specific time bins overlapping with MvMM time bins within the same neurons (fig. S8B).

To test whether the same population could represent multiple cross-modal identities, we developed the identity network model (INM), which integrates these two decoding approaches. The first approach was identical to the identity-specific decoder described above, which resulted in accurate decoding for each individual's face or voice. The second approach classified MvMM trials as either match or mismatch but was anonymized to individual identity. Our INM combined these two approaches to achieve cross-modal decoding of individual identity (fig. S9). This combination was critical because the identity-specific predictive population was only accurate for individual identity but performed poorly for classifying MvMM (Fig. 3G), whereas the MvMM predictive population was the opposite (Fig. 3H). When combined across individuals, the INM successfully decoded the cross-modal identity of all 12 individuals (accuracy: 84.5%) (Fig. 3I). Decoding performance tested at least 5× above chance when distinguishing all individuals (Fig. 3J and fig. S10).

Because identity neurons were included in decoding, we investigated whether their explanatory contribution was disproportionate to their sparse distribution. We compared INM performance when these neurons were removed from the analysis and separately used only in the analysis versus an equal number of other neurons. We observed no significant effect on decoding performance despite the consideration of only individuals preferred by identity neurons (Fig. 3K and figs. S11 and S12), which suggests that these highly selective neurons are no more notable for decoding the cross-modal identity of familiar individuals in the hippocampus than other neurons in the same population. Furthermore, no significant effect of identity neurons on decoding was demonstrated at a larger 5-SD response threshold.

Social category representations in hippocampus

An individual's identity is also coupled to their social relationships, such as their family. To test whether hippocampus encodes categorical attributes of social identity, we applied nonlinear dimensionality reduction techniques (*22*). Using mean firing rates consistent with studies of face and voice processing in the primate brain (*23*), we first verified that these reduction techniques were capable of separating the stimulus categories at multiple probe locations along the anterior-posterior axis (fig. S13). We next replicated the findings of the INM using the same identity-specific predictive time bins for marmoset faces and voices drawn from the entire hippocampal population and showed that manifold projections similarly separated individuals (Fig. 4A and fig. S14), including for different subpopulations of neurons (fig. S14G). This suggests that cross-modal identity representations are evident in the population activity of marmoset hippocampus.

To investigate whether representation of identity can be described by the relative timing of spikes, we computed manifold projections of spike times recorded during identity-match trials (Fig. 4B, left) using parameterless signed connection rate features. The signed connection rate from one neuron to another describes how they interact, which reveals statistical distributions specific to any given pair of neurons (Fig. 4B, right)—a facet of neural activity distinct from the firing rate of any single neuron. Each spike was concatenated into a feature vector, which was then projected onto the manifold as is shown by one symbol (Fig. 4C). The feature vector was computed as the signed connection rate to each neuron at each observation time. Each observation time was a spike time of the neuron with the greatest number of spikes. Results using this event-coded measure revealed excellent separability for identity-match trials (Fig. 4C), which thereby replicated the effect observed with the INM using a distinct facet of neural activity.

We next investigated whether social categories other than identity may likewise be represented in event-coded hippocampal activity. We tested whether representations of other marmosets' family members were distinct from nonfamily members for the two marmosets whose families were not included in the stimulus sets using two distinct quantifications of manifold projections, although the pattern was consistent for all subjects. First, results revealed a significant difference in the mean square range (MSR) of the manifold projections along this category boundary (Fig. 4D and fig. S15A), which suggests that a larger event-coded state space was occupied while observing family members (fig. S15B). Although these projections were supervised, the clustering that emerged on the basis of respective social relatedness was unsupervised. Second, we computed the unsupervised latent firing rate as the manifold projection of the absolute value of signed connection rate. Although individual identities did not separate (fig. S16A), we found trajectories that appeared stable in time and comparable across trials (fig. S16B). The motion of mean latent firing rate significantly separated social categories at multiple time points for all subjects (Fig. 4E and fig. S16, C and D). Together, these results demonstrate that neural representations of social identity in primate hippocampus are not only invariant to the sensory modality and comparable over time (fig. S17) but that low-dimensional manifolds (Fig. 4F) can describe relationships between different social categories (e.g., individual identity, family groups).

Conclusions

We showed that the cross-modal identity of multiple conspecifics is represented in the primate hippocampus. Although we identified putative concept cells similarly to human studies (*9*, *12*), we discovered that this population of highly selective neurons is not the only mechanism for representing concepts of individuals. Rather, both single neurons and the broader population in hippocampus encode cross-modal identity of multiple conspecifics, similar to what has been reported for objects (*24*), which suggests that the sparse representations of concept cells may not be the only mechanism to represent semantic memory in hippocampus. An important caveat to these findings, however, is that our criteria for determining the responsiveness of putative concept cells differed somewhat from previous studies in humans (*9*–*11*), as described above. It is possible, therefore, that the concept cells in marmoset and human hippocampus are not strictly analogous. Ultimately, whether these neurons are equivalent is not determined solely by the physiological properties used to classify them in analyses but their functional role in memory. To directly address this issue, comparative experiments examining the computational contributions of concept cells in hippocampus across species during memory are critical, as such data are currently lacking. In addition to these findings at the single-neuron level, a population-level code representing not only the cross-modal identity of multiple familiar individuals but information pertinent to social categories was likewise reported in this study. Information from both the putative concept cells and those neurons

that encoded multiple identities were integrated, which suggests that cross-modal identity in hippocampus is evident in the ensemble activity of this keystone brain structure. Similar to the role of hippocampus in other contexts (fig. S18) (*25*), the cross-modal identity representations revealed in this experiment may support a learned schema that here applies to social identity (*26*, *27*). That these experiments were performed in a highly constrained context limits our ability to determine whether such a schema would indeed be leveraged in more naturalistic contexts during which social decisions on the basis of conspecifics' identity are made continuously (*6*, *8*). The presence of unimodal representations of identity in the primate frontal and temporal cortex (*2*, *8*, *28*), amygdala (*5*, *29*), and the medial temporal lobe (*30*) and representations of social dominance in the amygdala (*31*) may reflect an integrative social recognition circuit in which substrates in the broader network play distinct but complementary roles that collectively govern natural primate social brain functions (*32*).

REFERENCES AND NOTES

1. R. M. Seyfarth, D. L. Cheney, in *The Evolution of Primate Societies*, J. Mitani, J. Call, P. M. Kappeler, R. Palombit, J. B. Silk, Eds. (Univ. Chicago Press, 2012), pp. 629–642.
2. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, *Curr. Biol.* **21**, 1408–1415 (2011).
3. P. Belin, C. Bodin, V. Aglieri, *Hear. Res.* **366**, 65–74 (2018).
4. J. Sliwa, A. Planté, J.-R. Duhamel, S. Wirth, *Cereb. Cortex* **26**, 950–966 (2016).
5. U. Rutishauser *et al.*, *Curr. Biol.* **21**, 1654–1660 (2011).
6. M. C. Rose, B. Styr, T. A. Schmid, J. E. Elie, M. M. Yartsev, *Science* **374**, eaba9584 (2021).
7. L. Chang, D. Y. Tsao, *Cell* **169**, 1013–1028.e14 (2017).
8. R. Báez-Mendoza, E. P. Mastrobattista, A. J. Wang, Z. M. Williams, *Science* **374**, eabb4149 (2021).
9. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, *Nature* **435**, 1102–1107 (2005).
10. R. Quian Quiroga, A. Kraskov, C. Koch, I. Fried, *Curr. Biol.* **19**, 1308–1313 (2009).
11. I. V. Viskontas, R. Q. Quiroga, I. Fried, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21329–21334 (2009).
12. R. Q. Quiroga, *Nat. Rev. Neurosci.* **13**, 587–597 (2012).
13. R. Quian Quiroga, *Hippocampus* **33**, 616–634 (2023).
14. H. S. Courellis *et al.*, *PLOS Biol.* **17**, e3000546 (2019).
15. F. Mormann *et al.*, *J. Neurosci.* **28**, 8865–8872 (2008).
16. J. Minxha *et al.*, *Cell Rep.* **18**, 878–891 (2017).
17. M. Rigotti *et al.*, *Nature* **497**, 585–590 (2013).
18. C. T. Miller *et al.*, *Curr. Biol.* **32**, R482–R493 (2022).
19. D. Kumaran, E. A. Maguire, *J. Neurosci.* **27**, 8517–8524 (2007).
20. M. Fyhn, S. Molden, S. Hollup, M.-B. Moser, E. Moser, *Neuron* **35**, 555–566 (2002).
21. T. Chen, C. Guestrin, in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016), pp. 785–794.
22. E. H. Nieh *et al.*, *Nature* **595**, 80–84 (2021).
23. W. A. Freiwald, D. Y. Tsao, *Science* **330**, 845–851 (2010).
24. T. P. Reber *et al.*, *PLOS Biol.* **17**, e3000290 (2019).
25. P. Baraduc, J.-R. Duhamel, S. Wirth, *Science* **363**, 635–639 (2019).
26. J. Sliwa, J.-R. Duhamel, O. Pascalis, S. Wirth, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1735–1740 (2011).
27. I. Adachi, R. R. Hampton, *PLOS ONE* **6**, e23345 (2011).
28. D. Y. Tsao, M. S. Livingstone, *Annu. Rev. Neurosci.* **31**, 411–437 (2008).
29. K. M. Gothard, F. P. Battaglia, C. A. Erickson, K. M. Spitler, D. G. Amaral, *J. Neurophysiol.* **97**, 1671–1683 (2007).
30. S. M. Landi, P. Viswanathan, S. Serene, W. A. Freiwald, *Science* **373**, 581–585 (2021).
31. J. Munuera, M. Rigotti, C. D. Salzman, *Nat. Neurosci.* **21**, 415–423 (2018).
32. W. A. Freiwald, *Curr. Opin. Neurobiol.* **65**, 49–58 (2020).
33. T. Tyree, M. Metke, C. Miller, Cross-modal representation of identity in primate hippocampus, Dataset, *Dryad* (2022).

ACKNOWLEDGMENTS

We thank H. Courellis for assistance with data collection and D. Leopold for comments on a previous version of this manuscript. **Funding:** This study was supported by National Institutes of Health grant R01 NS109294 (to C.T.M.) and National Institutes of Health grant R01 DC012087 (to C.T.M.). **Author contributions:** Conceptualization: C.T.M.; Methodology: T.J.T., M.M., and C.T.M.; Investigation: M.M. and T.J.T.; Analysis: T.J.T.; Visualization: T.J.T.; Funding acquisition: C.T.M.; Supervision: C.T.M.; Writing – original draft: T.J.T. and C.T.M.; Writing – review & editing: T.J.T. and C.T.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials or are deposited at Dryad (*33*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0460
Materials and Methods
Figs. S1 to S20
Table S1
References (*34*–*38*)
MDAR Reproducibility Checklist

Submitted 25 September 2022; resubmitted 30 May 2023
Accepted 1 September 2023
10.1126/science.adf0460

ENZYMOLOGY

Trivalent rare earth metal cofactors confer rapid NP-DNA polymerase activity

Victor S. Lelyveld[1,2], Ziyuan Fang[1,2,3], Jack W. Szostak[1,2,3,4,5,6]*

A DNA polymerase with a single mutation and a divalent calcium cofactor catalyzes the synthesis of unnatural N3′→P5′ phosphoramidate (NP) bonds to form NP-DNA. However, this template-directed phosphoryl transfer activity remains orders of magnitude slower than native phosphodiester synthesis. Here, we used time-resolved x-ray crystallography to show that NP-DNA synthesis proceeds with a single detectable calcium ion in the active site. Using insights from isotopic and elemental effects, we propose that one-metal-ion electrophilic substrate activation is inferior to the native two-metal-ion mechanism. We found that this deficiency in divalent activation could be ameliorated by trivalent rare earth and post–transition metal cations, substantially enhancing NP-DNA synthesis. Scandium(III), in particular, confers highly specific NP activity with kinetics enhanced by more than 100-fold over calcium(II), yielding NP-DNA strands up to 100 nucleotides in length.

The principal chemistry at the core of RNA and DNA metabolism is phosphodiester synthesis. Polymerases generate nascent strands of genetic material by stepwise phosphoryl transfer of nucleotides to primer termini, yielding O3′→P5′ phosphodiester linkages. This template-directed process is conserved across all known biology. Nucleotide 3′ substitutions have therefore been widely regarded as chain terminating for polymerase activity, forming the basis for a class of nucleoside analog drugs (*1*, *2*).

In recent work, we demonstrated the direct enzymatic synthesis of an unnatural linkage by substitution of the 3′-OH nucleophile with a 3′-amine, extending the chemistry amenable to polymerase catalysis (*1*). We reported that 3′-NH_2 primer extension can be catalyzed by a modified DNA polymerase, yielding N3′→P5′ phosphoramidate DNA (NP-DNA; Fig. 1, A and B). The large fragment of DNA polymerase I (BF), cloned from the thermophilic soil bacterium *Geobacillus stearothermophilus* (Bst), acquires nontrivial levels of N3′→P5′ polymerase activity through two unexpectedly minor substitutions: a single active-site mutation (F710Y) and substitution of its divalent Mg^{2+} cofactors with Ca^{2+} (*1*). However, this level of activity remained around four orders of magnitude slower than native phosphodiester synthesis. Although NP synthesis was detectable in the presence of several divalent alkaline earth metal ions, the pattern of metal cofactor activity was distinct from that found in native phosphodiester activity (*3*). Crystal structures of the ground-state NP reaction complex showed a single metal ion in the active site (*1*), suggesting a plausible distinction between the NP mechanism and the classical two-metal-ion mechanism in native phosphodiester activity (*4*). Whether NP catalysis does in fact rely on a distinct mechanism with a nonclassical cofactor configuration remains to be established, as do the critical factors limiting the kinetics of NP synthesis.

[1]Center for Computational and Integrative Biology, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. [4]Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA. [5]Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, MA 02114, USA. [6]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. Email: jwszostak@uchicago.edu

Linearly correlated density dynamics in reacting BF crystals during NP synthesis

To understand the mechanistic barriers to rapid enzymatic NP-DNA synthesis, we first sought to observe NP bond formation through crystallography. In previous x-ray crystal structures of the fully assembled NP polymerase preinsertion complex, we modeled a single divalent metal ion in the reaction center, occupying a site distal to the primer-terminal nucleophile (*1*). This site is equivalent to the so-called "B site" in the classical two-metal-ion reaction center and includes inner sphere ligands from the substrate triphosphate moiety, the side chains of Asp^{830} and Asp^{653}, and the backbone amide of Tyr^{654}. No evidence for an "A-site" metal proximal to the primer 3′-amino terminus was observed in these earlier prechemistry model structures, a finding consistent with the expectation that a neutral amine is the nucleophile in NP synthesis. However, earlier structures did not fully recapitulate the active NP reaction center or the product state (*1*). The A-site metal ion can also be poorly ordered even in structures of the native polymerase reaction complex. Nakamura *et al.* reported that accumulation of an additional metal ion in a "C site" of polymerase η, a polymerase Y family member, occurs in a manner linearly correlated with bond formation, suggesting that the product state's postchemistry metal configuration may be distinct from that observed in the prechemistry reaction complex (*5*). This observa… was subsequently extended to polymeras… family members (*6*, *7*). To add evidence that a single divalent metal ion cofactor catalyzes unnatural NP synthesis in BF, a polymerase A family member, we performed a single-nucleotide primer extension reaction in intact crystals (Fig. 1, B to D). We crystallized active substrate-bound BF polymerase F710Y/D598A with a 3′–amino-terminal DNA primer and DNA template in the presence of Ca^{2+} and 2′-deoxyguanosine 5′-triphosphate (dGTP). Reactions were then initiated in the intact crystals by a pH shift from an acidic mother liquor (pH < 6) to a basic soaking liquor at pH 8.8, and reacting crystals were quenched at various times by flash freezing (Fig. 1, C and D).

Quantitative analysis of electron density dynamics is complicated by crystal-to-crystal variance in diffraction datasets. We therefore incorporated datasets arising from up to four crystals at each of six time points after initiation. By scaling reflections from 19 total crystals (resolution 2.0 to 2.7 Å) across the time course (0 to 24 hours), we could estimate voxel-wise first-order rate constants for changes in real-space electron density across the asymmetric unit and particularly at the reaction center (Fig. 1D, bottom left panel). Early time points show negligible differences in ordered density between the ground-state complex crystallized under acidic conditions versus the complex after 1 hour of soaking under reaction conditions, suggesting that the reactant state and ground state cannot be clearly distinguished in this resolution range. By following the reaction at subsequent times, we observed monotonic accumulation of nascent NP bond density between the primer terminal 3′-amino group and the substrate α-phosphate with concomitant loss of electron density between the α- and β-phosphates with similar first-order rate constants (Fig. 1D).

By inspecting the kinetics of density changes at nearby sites, we observed that the most prominent local density dynamics in the active site had simple linear relationships with nascent bond formation. Pairwise linear regression yields the relative slope (β) of density changes occurring at any two points across the time-resolved dataset (Fig. 2, A and B). Performing this regression at all points in the asymmetric unit versus the nascent bond furnished a field of β values or "beta map" in real space, which could be contoured at a desired threshold of |β| to give isosurfaces for visualization (Fig. 2C). To show that regions with high |β| are also highly correlated, we produced pairwise Pearson correlation maps using a similar procedure (Fig. 2D). Regions of local conformational dynamics were qualitatively in agreement between correlation maps with coefficient r contoured at $|r| > 0.9$ (Fig. 2D) versus regression maps contoured at $|\beta| > 0.45$ (Fig. 2C). When beta maps were

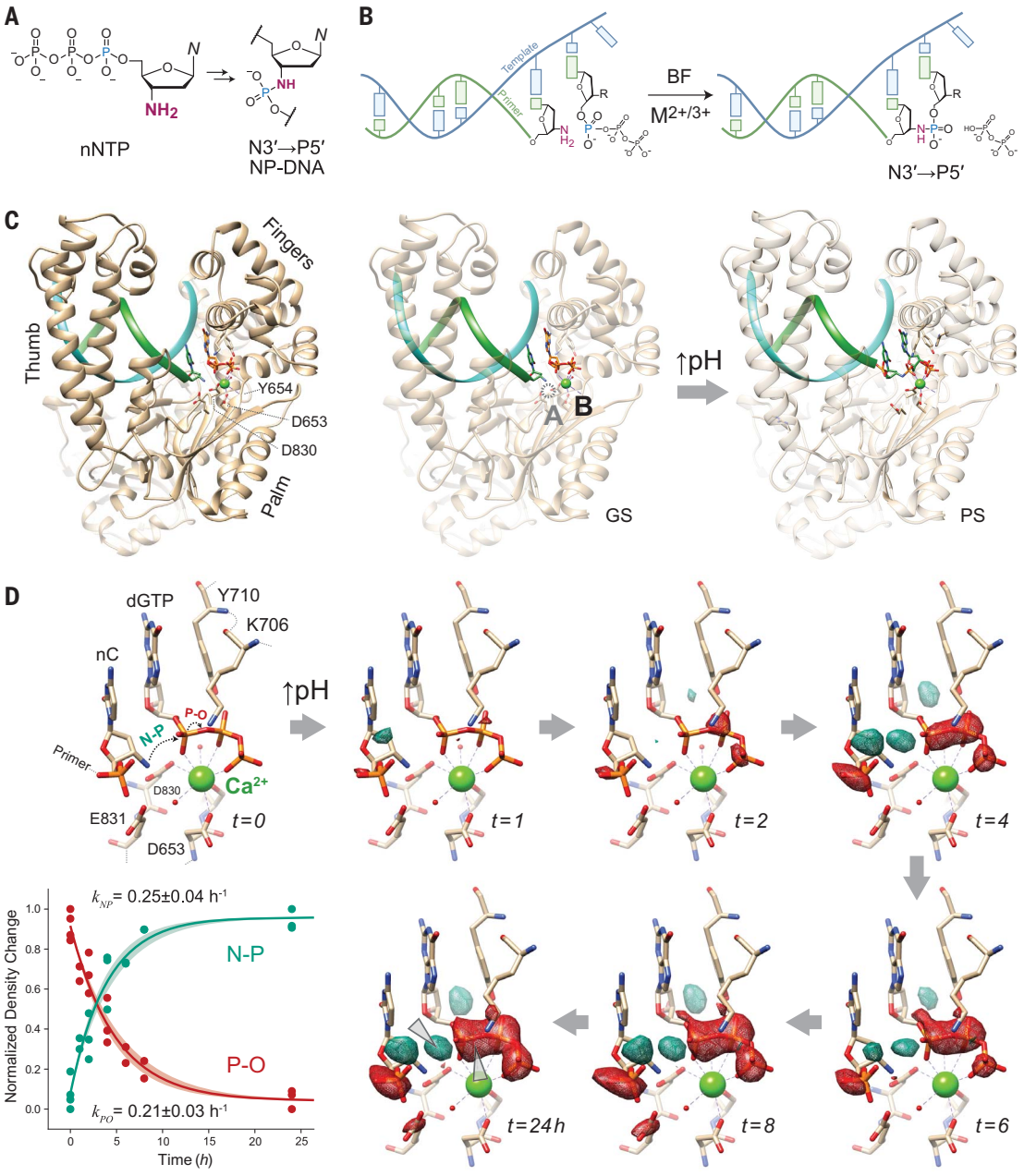


Fig. 1. NP polymerase activity proceeds in crystallo by a one-metal-ion mechanism. (**A**) Condensation of nNTPs generates NP-DNA. (**B**) Reaction scheme for BF-catalyzed primer extension of a 3′–amino-terminal primer on a DNA template, yielding pyrophosphate and an NP-linked +1 nucleotide extension. R = OH for in crystallo single-nucleotide addition of dGTP with co-crystallized divalent metal ion cofactor Ca^{2+}, whereas R = NH_2 for multiple turnover extension in solution to synthesize NP-DNA polynucleotide strands in the presence of divalent (M^{2+}) or trivalent (M^{3+}) metal ion cofactors. (**C**) Ribbon cartoons of BF polymerase ground-state (GS) holocomplex x-ray crystal structure with 3′–amino-terminal DNA primer (green ribbon) and DNA template (cyan) with bound dGTP substrate and a single Ca^{2+} ion (green atom) in the canonical metal-B site (left and middle panels), showing one of two complexes co-crystallized in the asymmetric unit. All analyses performed here derive from the first complex (structure chains A to C), but consistent trends are seen in both. Dotted circle indicates the position of a missing A-site metal in analogous models of the canonical phosphodiester synthesis mechanism but not observed in closed conformation GS structures containing a 3′-NH_2 (middle panel) or in product state (PS) structures of the in situ synthesized NP bond. (**D**) Representative time-dependent difference maps ($F_o - F_c$) showing density changes in the BF F710Y/D598A active site during NP bond formation in crystallo. GS crystals containing a 3′-amino-2′,3′-dideoxycytidine (nC) terminated primer and DNA template were formed under acidic conditions (pH 5 to 6) in the presence of Ca^{2+} and dGTP, and the in situ reaction was initiated by transferring crystals to a solution at pH 8.8. The reaction in intact crystals was quenched at the indicated times by flash freezing for subsequent data collection. Positive (aquamarine mesh) or negative (red mesh) isosurfaces are shown contoured at 2.5 σ and superimposed on the GS model. Bottom left inset: normalized density change extracted from difference maps at the nascent N–P (aquamarine) and scissile O–P (red) bonds quantified from 19 crystals (dots). Observed first-order rate constant estimates, k_{NP} and k_{PO}, for the density changes ± SD are plotted (solid line ± shaded area).

contoured at |β| > 0.9, we observed that the included voxels were tightly constrained to the site of the nascent NP bond and the scissile P_α–$O_{\alpha,\beta}$ bond (fig. S1A). No other linear density dynamics were larger in scalar magnitude than the chemistry itself during bond formation in crystallo.

Inspecting the isosurfaces contoured at |β| > 0.45, we observed four major density perturbations that were also highly correlated (|*r*| > 0.9) with the nascent bond (Fig. 2B, i to iv): (i) displacement of the terminal phosphodiester linkage between the –1 position and the primer terminus; (ii) a conformational switch of the substrate deoxyribose moiety from C2′-endo to C3′-endo, matching the sugar pucker of the primer terminus; (iii) a small deflection of the K706 side chain toward the substrate bridging α,β-oxygen, homologous to a putative general acid in the native mechanism (*8*); and finally (iv) disordering of the leaving group pyrophosphate moiety. As apparent from the high correlation coefficients, maps generated as 1 – *p* for two-tailed *P* values determined for the correlation field versus the nascent bond show that all of

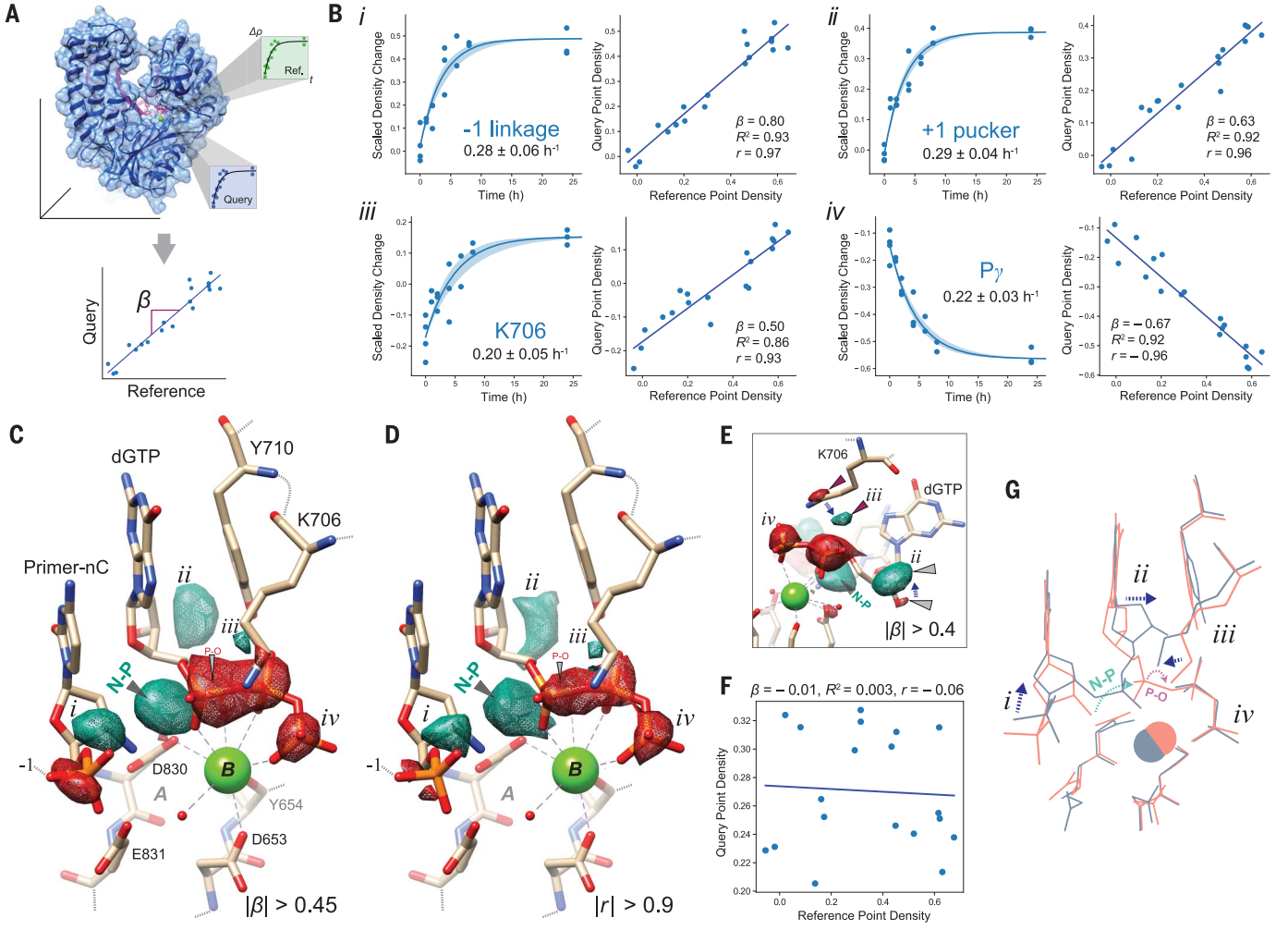


Fig. 2. One-metal-ion NP catalysis is limited by deficient substrate activation. (**A**) Voxelwise linear regression analysis of density dynamics in crystallo. Difference map ($F_o - F_c$) densities derived from reacting crystals quenched at various times yield kinetic traces at a reference point (Ref.) or any query point of interest. A pairwise linear regression slope (β) of density changes at the reference versus query points can be estimated for any two points in real space. (**B**) Regression statistics of key density dynamics during NP synthesis in BF F710Y/D598A. Insets *i* to *iv* show difference map density derived from individual crystals at four major active-site blobs indicated in (C). Left panels: first-order kinetics of density differences from individual crystals (dots) over reaction time, with values for the estimated rate constant ± SD plotted (solid line ± shaded area). Right panels: pairwise linear regressions, coefficients of determination (R^2), and correlation coefficients (*r*) of difference map densities at the indicated site versus the NP bond across all crystals. (**C**) Pairwise linear regression "beta" map calculated at all points versus the nascent NP bond (black wedge) with positive (aquamarine mesh) or negative (red mesh) β displayed as isosurfaces superimposed on the GS structure and contoured at |β| > 0.45. (**D**) Pearson correlation map of the active site contoured at |*r*| > 0.9 calculated by pairwise comparison of all voxels versus the peak of the nascent bond (black wedge), with positive (aquamarine mesh) or negative (red mesh) correlation displayed as isosurfaces superimposed on the GS structure. (**E**) Alternative view of the beta map contoured at |β| > 0.4, highlighting positive and negative beta peaks at sites (purple wedges) associated with a conformational change at the side chain of Lys^{706} [positive lobe quantified in (B), *iii*] in the direction of the substrate $O_{\alpha,\beta}$ bridging position and the scissile bond, as well as the conformation change in the dGTP substrate sugar [gray wedges, positive lobe quantified in (B), *ii*]. (**F**) Pairwise linear regression and correlation statistics versus the nascent bond calculated at a position equivalent to the absent A-site metal [gray "A" label in (C) and (D)]. (**G**) Cartoon overlay of GS (peach) and PS (blue) models with indicated conformational dynamics labeled as in (B) to (E).

these active-site dynamics were highly significant with $P < 10^{-6}$ (fig. S1B). Deflection of K706 toward the substrate α,β-bridging oxygen was well resolved on the beta map as positive and negative lobes surrounding terminal atoms of the side chain (Fig. 2E). This motion is noteworthy because the site is structurally homologous to the C-site metal-ion density reported for X- and Y-family polymerases (*5*, *6*, *7*, *9*). Crucially, no well-correlated density changes were observed at a position equivalent to the absent A-site metal in the native mechanism (Fig. 2F).

Conformational changes in both the upstream and downstream nucleotides flanking the nascent bond were linearly correlated with covalent bonding (Fig. 2B, i and ii and C and D). The upstream conformational change at the primer-terminal linkage was also observable in time-resolved datasets of native phosphodiester synthesis in polymerase η (*5*). The downstream conformational change in the dGTP substrate sugar from C2′- to C3′-endo upon product formation (Fig. 2, B, inset ii, and E), however, was not observed in an equivalent time-resolved experiment with the wild-type enzyme, in which the conformation was stably C3′-endo (fig. S2). Although interesting, a ground-state conformational change in the deoxynucleotide substrate is unlikely to be relevant to 3′-amino

nucleotides, which already prefer the C3′-endo conformation due to a mixture of steric and anomeric effects (see the supplementary materials) (*10*).

Isotopic and elemental substitution effects point toward rate-limiting chemistry with Ca^{2+}

Reacting crystals were sampled on a time scale vastly slower than that of conformational dynamics such as furanose pseudorotation. The crystal data are therefore not directly informative of the relative sequence and magnitude of these reaction barriers. To gain more insight into the rate-limiting step for NP synthesis, we turned to solution-phase kinetics. Two proton transfers have been detected in the rate-limiting step of native polymerase activity using careful measurements of the solvent deuterium kinetic isotope effect (SDKIE) (*8*, *11*). In NP synthesis, it is minimally required that one proton arising from the 3′-amino nucleophile must ultimately be transferred out of the reaction center in the forward reaction, yielding the product phosphoramidate. The conserved Asp^{830} side chain, proximal to the nucleophilic primer 3′-amine at ~2.5 Å in refined ground-state structures (Fig. 3B), is the most likely general base mediating this proton transfer. A sterically conservative point mutation at this position, D830N, entirely abolishes activity (*1*). We found that the effect of varying mole fraction, *n*, of D_2O solvent on pre–steady-state NP reaction kinetics was negligible, yielding an SDKIE estimate of 1.16 for 3′-amino primer extension in the presence of Ca^{2+} and dCTP (Fig. 3A). In native phosphodiester polymerases, this value has generally been measured in the 2 to 5 range (*8*). The substantially lower value measured here suggests that proton transfer is not a critical barrier in the NP reaction with Ca^{2+}.

If the chemical step of NP bond synthesis sets the overall reaction rate, then modulating the electrophilicity of the substrate should show an appreciable kinetic effect. We therefore measured the nonbridging elemental thio substitution effect in solution using stereopure *Sp* or *Rp* diastereomers of dCTPαS. We found that the thio effect was $k_O/k_S = 10.6 \pm 0.5$ (mean ± SD) when substituting the dCTP substrate with *Sp*-dCTPαS in pre–steady-state extension of a 3′-amino terminus at 45°C (Fig. 3, C and D). Elemental thio effects of similar magnitude have been observed for native polymerase mismatch extension or when the divalent cofactors are Mn^{2+} rather than Mg^{2+}, both conditions for which the chemical step appears to be rate limiting (*11*, *12*). The thio effect was compounded by an additional fourfold for the *Rp* substrate versus the *Sp* substrate (Fig. 3D), consistent with the coordination geometry seen in crystal structures (Fig. 3B). The interpretation of thio effects has long been a matter of debate in phosphoryl transfer catalysis, given the potential for complicating

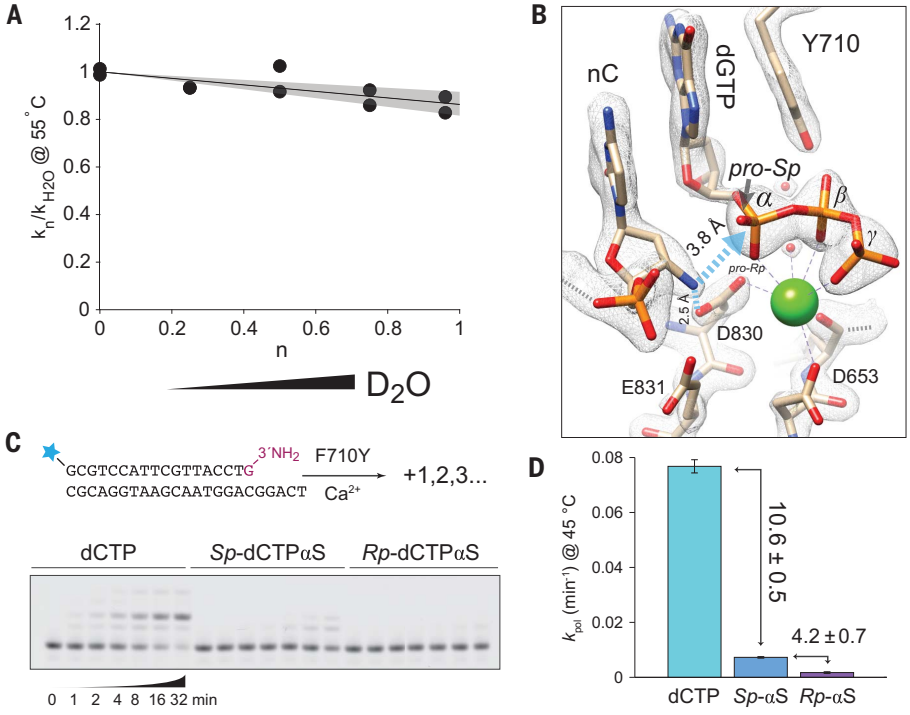


Fig. 3. Probing for rate-limiting chemistry in NP-DNA synthesis in solution. (**A**) Pre–steady-state solvent deuterium kinetic isotope effect estimated from the slope of the line $k_n/k_{H2O} = 1 + n(\varphi\text{-}1)$ for 3′-amino primer extension with 1 mM dCTP and 10 mM $CaCl_2$ in varying mole fractions, *n*, of D_2O. The quantity $1/\varphi$ is the SDKIE, estimated at 1.16 with 95% confidence interval (1.09 to 1.23) in the shaded region. (**B**) Atomic model distances in the GS reaction complex for the primer-terminal (nC) 3′-amine to the D830 side chain or to the substrate P_α. Gray mesh overlay indicates $2F_o - F_c$ density map contoured at 2.5 σ. (**C**) Effect of α-phosphorothioate substitution on pre–steady-state Ca^{2+}-activated NP synthesis at 45°C with 500 μM *Sp*-αS, *Rp*-αS, or unmodified dCTP substrates and BF F710Y. Inset: fluorescently labeled primer and template strand for the experiments in (A), (C), and (D). Representative 15% tris-borate EDTA (TBE)–urea PAGE separation of quenched reaction samples is shown. Only the first addition to this primer forms an NP or NPS linkage in the presence of 2′-deoxyribose substrates, yielding subsequent phosphodiester or phosphorothioate products, including mismatch extension products. (**D**) Elemental thio effects (labeled arrows, mean ± SD) estimated from the pre–steady-state rate constants (k_{pol}, error bars indicate SD, *n* = 3) for reactions with substrates as in (C).

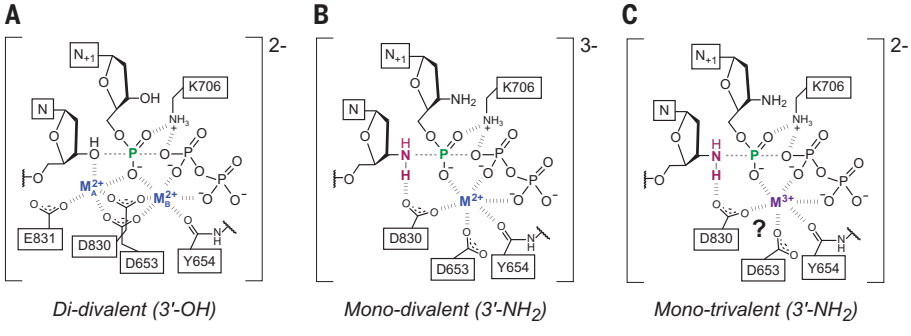


Fig. 4. Comparison of reaction models for phosphodiester and phosphoramidate synthesis in BF. (**A**) Model for the canonical two-metal ion "di-divalent" mechanism for native phosphodiester activity. (**B**) Model for the mono-divalent mechanism for Ca^{2+}-catalyzed NP activity. (**C**) Proposed mono-trivalent mechanism for REE-catalyzed NP activity. Note that neither inner sphere waters nor exact proton-transfer pathways have been depicted. Two rate-limiting proton transfers have been detected for native activity (*8*, *11*), but proton transfer is not rate limiting for mono-divalent NP activity with Ca^{2+}. At least one proton must nevertheless be transferred to generate the phosphoramidate product from the attacking neutral amine. In BF, these transfers likely involve deprotonated D830 as a general base and protonated K706 as a general acid.

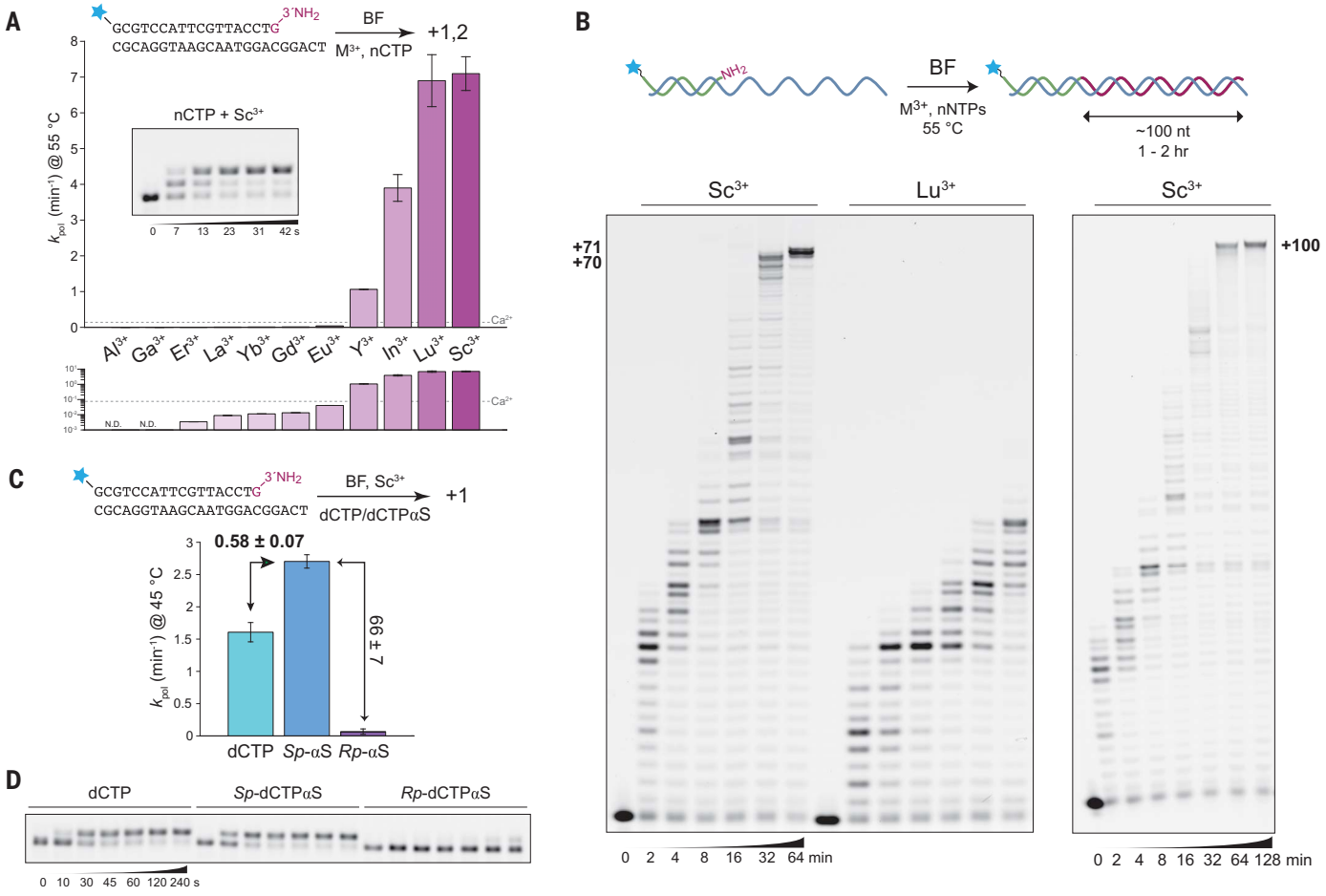


Fig. 5. Trivalent rare earth metal ion cofactors confer rapid NP polymerase activity. (**A**) Pre–steady-state k_{pol} estimates for extension of a 3′–amino-terminal DNA primer on a DNA template (inset cartoon) by BF F710Y/D598A at 55°C in the presence of various 1:1 ammonium citrate–buffered trivalent metal cations at 5 mM in 40 mM Tris-HCl, pH 8.8, 2 mM βME, in reactions initiated by addition of 250 μM nCTP. Estimates are displayed in linear (top) and log-scaled (bottom) axes compared with the level of Ca^{2+} activity (dotted line). Inset: representative 15% TBE-urea PAGE of quenched reaction samples from BF F710Y/D598A 3′–amino primer extension reactions in the presence of 5 mM Sc^{3+} showing +1 and +2 NP-DNA extension at the indicated times. ND, not detected. Error bars indicate SEM for n = 4 (Sc^{3+} and Lu^{3+}) or n = 3 (In^{3+} and Y^{3+}). (**B**) Mixed-sequence NP-DNA synthesis with excess BF F710Y/D598A on a 71-nt (left image) or 100-nt (right image) DNA-templating region at 55°C from a 5′–fluorescein-labeled 3′–amino-terminal DNA primer in the presence of 1:1 ammonium citrate–buffered 5 mM Sc^{3+} or Lu^{3+}, as indicated, in 40 mM Tris-HCl, pH 8.8, 10 mM βME, and 25 μM spermine-HCl. Reactions were initiated by the addition of a 250 μM concentration of each nNTP, sampled and quenched at the indicated times, and samples were separated on 10% (left) or 8% (right) TBE-urea PAGE gels. (**C**) Elemental thio effects (labeled arrows, mean ± SD) for Sc^{3+}-activated NP synthesis with 500 μM *Sp*-αS, *Rp*-αS, or unmodified dCTP substrates at 45°C estimated from k_{pol} (error bars indicate SD, n = 3) using a fluorescently labeled primer and template strand (cartoon at top), yielding exclusively +1 product due to high NP(S) specificity. (**D**) Representative 15% TBE-urea PAGE separation of quenched samples for the reactions in (C).

steric effects with phosphorothioate substrates (*12*, *13*). Nevertheless, a plausible interpretation of the observed thio effects is that the chemical step sets the overall reaction rate for NP synthesis.

Electrophilic substrate activation as a probable deficiency in NP catalysis with Ca^{2+}

If the chemical step is in fact rate limiting but proton transfer is not, then an augmented barrier to NP versus phosphodiester synthesis could, at least in part, originate from an altered metal-ion configuration in the presence of a 3′-amino primer. In the native reaction center, it is generally understood that the A-site metal ion activates the 3′-OH nucleophile by inner sphere coordination, yielding a metal alkoxide in the transition state, but numerous crystallographic models of reaction intermediates also show that the octahedral ion bound at the A site forms an additional inner sphere contact with the substrate pro-*Rp* α-phosphate nonbridging oxygen (*14*). It has been independently argued that the role of the conserved polymerase dinuclear metal center (Fig. 4A) is to stabilize the transition state electrostatically on the basis of linear free energy relationships obtained from Brønsted plots of pre–steady-state kinetics in polymerase β, harnessing a series of $O_{\beta,\gamma}$ bridge substituent–modified substrate analogs with varying leaving group pK_a (*15*, *16*). The catalytic effect of the A-site metal in native activity likely encompasses several effects, but any catalytic effect conferred by the presence of an A-site metal ion on substrate activation, as opposed to nucleophile activation, is absent for NP synthesis. The lack of electron density for an A-site metal ion in the ground state and at later time points is also consistent with the expectation of a neutral nucleophile in the NP reaction and the generally weak affinity of aliphatic amines for Ca^{2+}. Because neither the apparent binding constant nor the active-site conformation is significantly perturbed by the presence

of a 3′-amino nucleophile (*1*), a significant contributor to the kinetic defect for NP versus phosphodiester synthesis may therefore be the relative loss of transition state stabilization associated with the missing A-site metal cofactor for the reaction with an amino nucleophile.

If the native metal cofactors act, at least in part, on the native transition state by charge stabilization, then it is noteworthy that the net effect of the absence of one divalent ion (with disengagement of its conserved ligand, Glu^{831}) would be to decrease the formal net charge of the transition state complex by 1 (Fig. 4, A and B). Outer sphere charge mutations may not be able to compensate for this inner sphere defect, particularly if the role of the missing metal in native activity is partially to contribute to stabilizing the developing charge polarization across the scissile bond. This view is consistent with the observation that eliminating the negative charge on the disengaged native ligand by the mutation E831Q fails to enhance NP synthesis kinetics (*1*).

Trivalent metal ions confer rapid and specific NP synthesis activity

On the basis of this activation argument and the observed thio effect, we hypothesized that substitution of a trivalent cation into the metal site B might compensate for the electrostatic component of the catalytic defect resulting from the absence of the A-site metal, following the hypothetical reaction structure depicted in Fig. 4C. Upon screening a series of redox-stable trivalent metal ions, we indeed found that an exotic series of trivalent metal ions could act as polymerase cofactors for rapid catalysis of NP bond formation (Fig. 5). The diamagnetic group three trivalent rare earth element (REE) cations scandium (Sc^{3+}), yttrium (Y^{3+}), and lutetium (Lu^{3+}), as well as the post–transition metal ion indium (In^{3+}), all significantly improved single-nucleotide 3′-amino primer extension (Fig. 5A), as well as NP-DNA polymerization with all four 3′-amino 2′,3′-dideoxynucleoside 5′-triphosphates (nNTPs) on mixed-sequence templates (Fig. 5B and figs. S3 and S4). Sc^{3+} in particular accelerated pre–steady-state rate constants for nCTP addition by ~100-fold at 55°C to 7.1 ± 0.5 min^{-1} (mean ± SEM) versus 0.069 min^{-1} for Ca^{2+} (*1*), suggesting a stabilization effect of approximately –3 kcal/mol. Lu^{3+} yielded similar levels of burst activity, k_{pol} = 6.9 ± 0.7 min^{-1} (mean ± SEM), but had far weaker multiple turnover activity in long primer extensions relative to Sc^{3+} (Fig. 5B and fig. S3). NP-active trivalent ion cofactors have a wide range of pK_a values for the aquo ion, are prone to hydrolysis under the reaction conditions, and have a tendency to precipitate nucleotides and other phosphates. We found that reaction conditions were optimal when these ions were buffered at ~1:1 stoichiometry with citrate such that all reaction components remain homogeneously soluble (fig. S4). Although the reaction kinetics were quite sensitive to the metal:citrate stoichiometry, they were not appreciably sensitive to the concentration of 1:1 Sc^{3+}:citrate in the low-millimolar range (0.5 to 10 mM) in the presence of 1 mM total nNTPs (fig. S4). However, it is known that the metal-ligand stoichiometry of trivalent REE citrate complexes in solution is diverse (*17*).

Trivalent metal ion cofactors confer exquisitely specific catalysis of NP versus phosphodiester chemistry. We found that the pre–steady-state rate constant for extension of a native DNA 3′-OH terminus with nCTP in the presence of Sc^{3+} was 5.5 ± 0.6 × 10^{-3} min^{-1}, which is at least three orders of magnitude slower than 3′-NH_2 extension (fig. S3B). NP-DNA synthesis activated by Sc^{3+} was also inhibited by added Mg^{2+} in the low millimolar range optimal for native activity (fig. S5A). The mutation D830N was completely inactivating, whereas the adjacent E831Q mutant exhibited wild-type activity in long extensions (fig. S5B), consistent with the active-site configuration proposed in Fig. 4C. Another notable feature of trivalent NP catalysis is an apparent inverse elemental thio effect. We found that the Sc^{3+} thio effect was 0.58 ± 0.07 (mean ± SD) at 45°C, indicating a significant preference for phosphorothioate substrates to yield thiophosphoramidate (NPS) linkages (Fig. 5, C and D). Inverse thio effects are more readily explained in the context of inverted stereospecificity with thiophilic metal cofactor substitution, but in this case the catalyzed reaction remained highly specific (66-fold) for the *Sp* over *Rp* substrates at 45°C. Synthesis of the potentially clinically relevant NPS linkages under these conditions was confirmed by high-resolution mass spectrometry (fig. S6).

NP-DNA synthesis on long templates was aided by the addition of polyamines (fig. S4). With spermine and Sc^{3+}, full-length products on mixed-sequence DNA templates up to +71 nucleotides (nt) or, with additional enzyme, up to +100 nt, could be synthesized in 1 to 2 hours at 55°C (Fig. 5B and fig. S4). Polyamine rescue is consistent with linkage-dependent conformational effects on duplex helicity for NP-DNA:DNA hybrids that arise during long-range primer extension synthesis, or it might reflect competition for inhibitory metal coordination sites. Activation by spermine was optimal at low micromolar concentrations, consistent with its reported effects on Klenow activity previously attributed to helicity effects (*18*). Under these conditions, NP-DNA synthesis remains highly determined by template sequence, because omission of any single aminonucleotide from the nNTP mixture markedly stalled extension activity at or immediately before the first template position complementary to the missing substrate (fig. S7A and B). Extension products also show characteristic features expected from the incorporation of unnatural NP linkages, such as acid lability at elevated temperature (fig. S7A and B, HOAc lanes) and high resistance to 3′-5′ exonuclease I activity (fig. S7, C and D).

Discussion

Unlike native phosphodiester catalysis, only a single metal ion cofactor is detectable by crystallography during NP catalysis. Although transition states and other short-lived intermediates were not directly observable in our studies, the totality of the evidence supports a single–metal ion mechanism with weakened substrate activation in 3′-amino primer extension by BF. From this inference, we identified a series of trivalent metal ion polymerase cofactors with markedly enhanced reactivity and specificity for NP-DNA synthesis. Relative to the wild-type BF DNA polymerase, combining a single active-site point mutation with trivalent REE cofactor substitution accelerated NP synthesis by well over 1000-fold, yielding template-directed NP-DNA strands of useful lengths in benchtop enzyme reactions. In the presence of Sc^{3+} and spermine, overall NP polymerase reaction time to yield full-length products (~0.9 to 1.3 min/nt at 55°C) was similar to that reported for certain evolved xenonucleic acid (XNA) polymerases producing OP-linked strands incorporating synthetic sugars [0.75 to 1.5 min/nt at 50 to 65°C (*19*)]. REE-catalyzed NP polymerase reactions are therefore comparable to those of extensively mutated polymerases engineered for alternative phosphodiester synthesis activities by directed evolution (*20*). This finding suggests that the evolutionary distance to catalysis of distinct chemistry may be negligible when alternative cofactors are present, given a conservatively modified substrate. By comparison, recovering a conserved chemistry with highly divergent substrates may require a more extensive search of sequence space.

The considerable catalytic advantage conferred by trivalent metal ion cofactors implies that a broader family of polymerase activities may be accessible by directed evolution of substrate specificity given the well-demonstrated ability of evolutionary methods to optimize native catalysis on XNA substrates (*20*, *21*). NP-DNA falls within a class of phosphoramidate nucleic acids that have long been valued for their potential utility as nuclease-resistant antisense oligonucleotides (*22*). These polymers have also been extensively studied for their nonenzymatic self-assembly from chemically activated phosphorimidazolide nucleotide analogs (*23*–*26*) on a path to developing synthetic protocells and model systems for key aspects of abiogenesis (*27*, *28*). Despite lacking a 2′-OH group, short synthetic NP-DNA duplexes adopt a conformational geometry more similar to RNA than to DNA (*10*) such that

these polymers are viable templates for reverse transcriptase activity (*29*) but are poor substrates for several classes of nucleases (*30*, *31*). Whether NP polymers are, like RNA, fully functional Darwinian polymers remains a matter of considerable interest. The development of practical levels of NP polymerase activity is a crucial step on the path to demonstrating that this class of synthetic genetic materials is in fact functional and evolvable.

REFERENCES AND NOTES

1. V. S. Lelyveld, W. Zhang, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7276–7283 (2020).
2. T. S. Mansour, R. Storer, *Curr. Pharm. Des.* **3**, 227–264 (1997).
3. A. K. Vashishtha, J. Wang, W. H. Konigsberg, *J. Biol. Chem.* **291**, 20869–20875 (2016).
4. T. A. Steitz, *Nature* **391**, 231–232 (1998).
5. T. Nakamura, Y. Zhao, Y. Yamagata, Y. J. Hua, W. Yang, *Nature* **487**, 196–201 (2012).
6. B. D. Freudenthal, W. A. Beard, D. D. Shock, S. H. Wilson, *Cell* **154**, 157–168 (2013).
7. J. A. Jamsen *et al*., *Nat. Commun.* **8**, 253 (2017).
8. C. Castro *et al*., *Nat. Struct. Mol. Biol.* **16**, 212–218 (2009).
9. Y. Gao, W. Yang, *Science* **352**, 1334–1337 (2016).
10. V. Tereshko, S. Gryaznov, M. Egli, *J. Am. Chem. Soc.* **120**, 269–283 (1998).
11. C. Castro *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4267–4272 (2007).
12. C. M. Joyce, S. J. Benkovic, *Biochemistry* **43**, 14317–14324 (2004).
13. D. Herschlag, J. A. Piccirilli, T. R. Cech, *Biochemistry* **30**, 4844–4854 (1991).
14. W.-J. Wu, W. Yang, M.-D. Tsai, *Nat. Rev. Chem.* **1**, 0068 (2017).
15. C. A. Sucato *et al*., *Biochemistry* **47**, 870–879 (2008).
16. K. Oertell *et al*., *Biochemistry* **53**, 1842–1848 (2014).
17. S. A. Wood, *Eng. Geol.* **34**, 229–259 (1993).
18. A. Osland, K. Kleppe, *Biochim. Biophys. Acta* **520**, 317–330 (1978).
19. A. I. Taylor, P. Holliger, *Nat. Protoc.* **10**, 1625–1642 (2015).
20. N. Freund, M. J. L. J. Fürst, P. Holliger, *Curr. Opin. Biotechnol.* **74**, 129–136 (2022).
21. A. A. Henry, F. E. Romesberg, *Curr. Opin. Biotechnol.* **16**, 370–377 (2005).
22. S. Gryaznov *et al*., *Nucleic Acids Res.* **24**, 1508–1514 (1996).
23. R. Lohrmann, L. E. Orgel, *Nature* **261**, 342–344 (1976).
24. W. S. Zielinski, L. E. Orgel, *Nucleic Acids Res.* **13**, 2469–2484 (1985).
25. S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17732–17737 (2013).
26. W. Zhang, A. Pal, A. Ricardo, J. W. Szostak, *J. Am. Chem. Soc.* **141**, 12159–12166 (2019).
27. S. S. Mansy *et al*., *Nature* **454**, 122–125 (2008).
28. D. K. O'Flaherty, L. Zhou, J. W. Szostak, *J. Am. Chem. Soc.* **141**, 10481–10488 (2019).
29. V. S. Lelyveld, D. K. O'Flaherty, L. Zhou, E. C. Izgu, J. W. Szostak, *Nucleic Acids Res.* **47**, 8941–8949 (2019).
30. S. Gryaznov, J.-K. Chen, *J. Am. Chem. Soc.* **116**, 3143–3144 (1994).
31. R. L. Hanna, S. M. Gryaznov, J. A. Doudna, *Chem. Biol.* **7**, 845–854 (2000).

ACKNOWLEDGMENTS

We thank the staff at the Advanced Light Source (ALS) beamline 822 and the Advanced Photon Source (APS) beamline 23ID-B for technical assistance and F. Ng for invaluable research support. **Funding:** This work was supported by the Howard Hughes Medical Institute and Simons Foundation (grant 290363 to J.W.S.). The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The ALS is a Department of Energy (DOE) Office of Science User Facility supported by contract no. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences (grant P30 GM124169). GM/CA@APS has been funded by the National Cancer Institute (grant ACB-12002) and the National Institute of General Medical Sciences (grant AGM-12006, P30GM138396). This research used resources of the APS, a DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. **Author contributions:** V.S.L. and J.W.S. conceived the study and designed the research plan. Z.F. performed crystallizations, data collections, and refinements. V.S.L. analyzed the structural data, prepared materials, performed solution phase experiments, and analyzed the data. V.S.L. and J.W.S. interpreted the data and wrote the manuscript. **Competing interests:** A related patent application regarding enzymatic polymerization of NP-linked polynucleotides has been filed by V.S.L. and J.W.S. The remaining author declares no competing interests. **Data and materials availability:** Crystallographic data and maps have been deposited into the Protein Data Bank under PDB codes 8SCG, 8SCI, 8SCJ, 8SCK, 8SCL, 8SCM, 8SCN, 8SCO, 8SCP, 8SCQ, 8SCR, 8SCS, 8SCT, and 8SCU. All other data are presented in the main text or supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the Author Accepted Manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh5339
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S3
References (*32*–*39*)
MDAR Reproducibility Checklist

Submitted 23 April 2023; accepted 20 September 2023
10.1126/science.adh5339

SPINTRONICS

Observation and control of hybrid spin-wave–Meissner-current transport modes

M. Borst, P. H. Vree, A. Lowther, A. Teepe, S. Kurdi, I. Bertelli, B. G. Simon, Y. M. Blanter, T. van der Sar*

Superconductors are materials with zero electrical resistivity and the ability to expel magnetic fields, which is known as the Meissner effect. Their dissipationless diamagnetic response is central to magnetic levitation and circuits such as quantum interference devices. In this work, we used superconducting diamagnetism to shape the magnetic environment governing the transport of spin waves—collective spin excitations in magnets that are promising on-chip signal carriers—in a thin-film magnet. Using diamond-based magnetic imaging, we observed hybridized spin-wave–Meissner-current transport modes with strongly altered, temperature-tunable wavelengths and then demonstrated local control of spin-wave refraction using a focused laser. Our results demonstrate the versatility of superconductor-manipulated spin-wave transport and have potential applications in spin-wave gratings, filters, crystals, and cavities.

The ability to control the transport of spins and charges with metal electrodes is fundamental to information-processing devices and an indispensable tool in quantum and condensed-matter physics. Although devices such as spin valves and transistors are based on the transport of uncorrelated particles (*1*), the excitations of magnetic materials, known as spin waves, are emerging as promising alternative information carriers (*2*). These collective spin excitations provide new opportunities for realizing analog or binary device functionality based on their wave nature, nonreciprocal transport properties, and low intrinsic damping (*3*).

Control of spin-wave transport is possible by heavy-metal electrodes that enable modulation by means of the spin-Hall effect (*4*–*6*) or by auxiliary magnetic materials that modify the spin-wave spectrum (*7*, *8*). However, metallic gates can also introduce additional spin-wave damping because of uncontrolled spin pumping or spin-wave–induced eddy currents (*6*, *9*, *10*). Furthermore, the diamagnetic response of normal metals is dominated by ohmic resistance, precluding effective stray-field control of the spin-wave spectrum.

An attractive approach for strong, low-damping spin-wave modulation is to use superconducting electrodes. Superconductors are materials with zero electrical resistivity and a strong diamagnetic response that enables the creation of magnetic shields, magnetic lenses, and circuits such as quantum bits and quantum interference devices (*11*, *12*). Spin-wave spectroscopy measurements have demonstrated that superconducting strips on magnetic films can alter the spin-wave spectrum through the backaction of induced currents (*13*) or the interaction with Abrikosov vortices (*14*). Recently, it was proposed to harness the diamagnetism of a superconductor to create the spin-wave equivalents of optical mirrors and cavities (*15*). Being able to image and control spin waves as they travel underneath superconducting electrodes would enable insight into the nature of the spin wave–superconductor interaction and unlock opportunities to control the propagation, dispersion, and refraction of spin waves.

In this work, we developed, imaged, and studied temperature-, field-, and laser-tunable spin-wave transport enabled by a superconducting strip on a thin-film magnetic insulator (Fig. 1A). We used magnetic resonance imaging based on nitrogen-vacancy (NV) spins in diamond (*16*–*18*) to study the spin waves as they travelled underneath the optically opaque superconductor.

Imaging hybrid spin-wave–Meissner-current transport modes using spins in diamond

Our system consisted of a thin film of yttrium iron garnet (YIG), a magnetic insulator with low spin-wave damping (*2*), equipped with gold microstrips for spin-wave excitation and a molybdenum-rhenium superconducting strip for spin-wave modulation (Fig. 1A). To image the spin waves, we placed a diamond membrane that contained a thin layer of NV sensor spins on top of the sample (Fig. 1A and fig. S1) (*18, 19*). These spins detect the spin waves by their microwave magnetic stray fields, which enables imaging through optically opaque materials (*10*). The sample was embedded in a variable temperature cryostat with a base temperature of 5.5 K and free-space optical access to read out the NV sensor spins.

NV centers are atomic defects in the diamond carbon lattice with an $S = 1$ electron spin (*16*). The sensitivity of the NV spin to magnetic fields, combined with its optical spin readout and excellent spin coherence, has enabled widespread sensing applications in fi ranging from condensed-matter science to ology and biophysics (*20*–*22*). Here we used the sensitivity of the NV spins to microwave magnetic fields to image the spin waves in our YIG film (*17*, *18*, *23*, *24*). When resonant with an NV electron spin resonance (ESR) frequency, the stray field of the spin waves drives transitions between the NV spin states, which we detected through the spin-dependent NV photoluminescence under green-laser excitation.

We applied a magnetic bias field to tune the NV ESR frequency into resonance with spin waves of different spin wavelengths (Fig. 1B). By orienting the field along one of the four possible crystallographic NV orientations (Fig. 1A), we split the ESR frequency of this "field-aligned" NV ensemble (f_1) off from that of the three other NV ensembles (f_2) as shown in the optically detected resonance spectrum of Fig. 1C. Alternatively, we applied the field in plane along $\hat{\boldsymbol{z}}$ to enable measurements at different frequencies at a given magnetic field. Because the applied magnetic fields are much smaller than the YIG saturation magnetization, the YIG magnetization lies predominantly in-plane along $\hat{\boldsymbol{z}}$ for both field orientations (Damon-Eshbach geometry).

To demonstrate the spin-wave modulation capabilities of our superconductor, we imaged the spin-wave transport above and below the molybdenum-rhenium (MoRe) superconducting transition temperature $T_c = 8.7$ K (Fig. 2, A to D). We generated NV-resonant spin waves with wavevector $\boldsymbol{k} = k\hat{\boldsymbol{y}}$ by applying a microwave current at NV frequency f_1 to the gold microstrip located just left outside of the imaging area. The interference between the microwave magnetic stray field generated by these spin waves and the direct microstrip field leads to a spatial standing-wave modulation of the NV ESR contrast (*17*, *18*). Crucial for our measurements, this interference effect enables a straightforward extraction of the spin wavelength. The spatial map of the ESR contrast C_1 at $T = 10.7$ K (above T_c) shows spin waves traveling toward and then underneath the MoRe strip without a change in wavelength (Fig. 2B). By contrast, the spin-wavelength increased almost twofold when the strip was cooled into its superconducting state at $T = 5.5$ K (Fig. 2C). Averaging the maps along $\hat{\boldsymbol{z}}$ (Fig. 2E) highlights the spatial homogeneity of the wavelength change.

We explained the superconductor-induced change of the spin wavelength by developing an analytical expression for the spin-wave dispersion in a magnet-superconductor thin-film hybrid. In this model, building on the formalism developed in (*15*), the spin waves induce AC Meissner currents that are governed by the London penetration depth λ_L of the superconductor. These currents, in turn, generate a magnetic field that acts back on the spin waves.

Department of Quantum Nanoscience, Kavli Institute of Nanoscience, Delft University of Technology, 2628 CJ Delft, Netherlands.
*Corresponding author. Email: t.vandersar@tudelft.nl

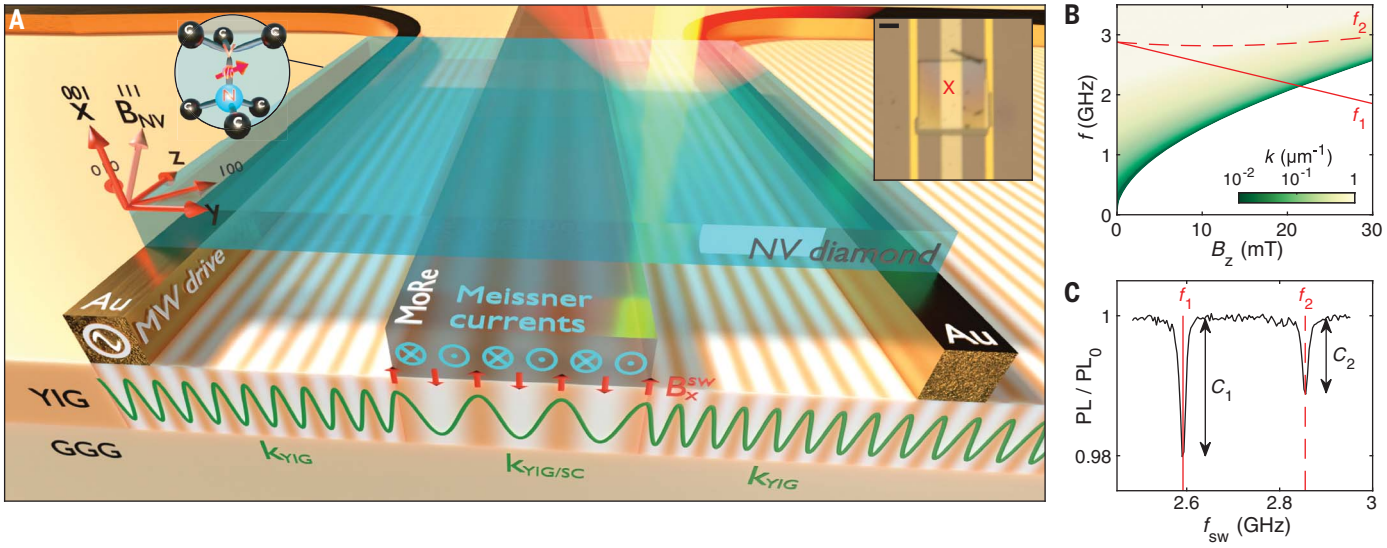
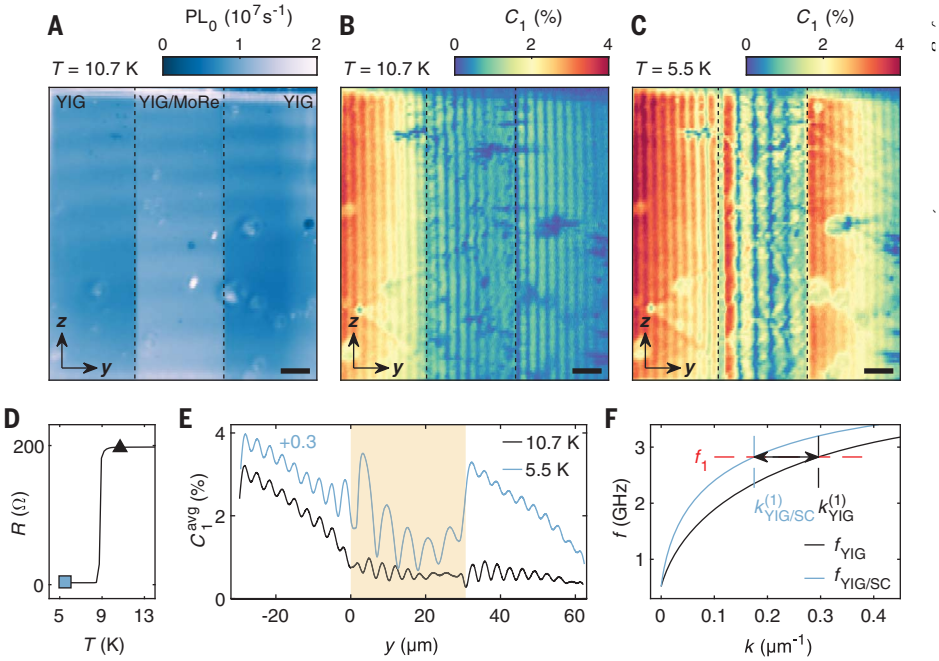


Fig. 1. Magnetic resonance imaging of hybridized spin-wave–Meissner-current transport modes. (**A**) Overview of the experiment. A gold (Au) microstrip excites spin waves with wavevector $\boldsymbol{k} = k\hat{\boldsymbol{y}}$ in a 245-nm-thick film of yttrium iron garnet (YIG). The spin waves travel toward a molybdenum rhenium (MoRe) superconducting strip (width W = 30 μm, thickness t = 140 nm) where their stray fields induce Meissner currents that act back on the spin waves, shifting their wave number to $k_{YIG/SC} < k_{YIG}$ (SC, superconductor). We imaged the waves underneath and next to the superconductor by their microwave magnetic stray fields using a ~20-nm-thick layer of nitrogen-vacancy (NV) spins implanted at ~0.07 μm below the bottom surface of a 100 × 100 × 5 μm^3 diamond membrane placed on top of the sample. A magnetic field $\boldsymbol{B}_{NV} = B_x\hat{\boldsymbol{x}} + B_z\hat{\boldsymbol{z}}$ applied at θ = 54° with respect to the x axis yields an in-plane YIG magnetization along $\hat{\boldsymbol{z}}$ for the small fields applied and directional Damon-Eshbach spin-wave excitation. GGG, gadolinium gallium garnet, SW, spin wave. (Inset) Optical micrograph of NV diamond and Au and MoRe strips on the YIG. Scale bar, 30 μm. Au thickness, 200 nm. (**B**) YIG dispersion (color map) and NV electron spin resonance (ESR) frequencies (red lines) as a function of the in-plane field component $B_z = B_{NV}\cos\theta$. f_1 and f_2 denote the ESR frequencies of NV spins with zero-field quantization axis aligned or misaligned with $\boldsymbol{B}_{NV}$, respectively. The intersection of the ESR frequencies with the spin-wave dispersion sets the detectable spin-wave numbers k. (**C**) Optically detected NV ESR spectrum at B_z = 10 mT, at a location denoted by the red cross in the inset of (A). The ESR contrast C_1 or C_2, where the subscript identifies the field-aligned or misaligned NV ensemble, respectively, results from interference between the microstrip field and spin-wave field, enabling spatial mapping of the spin-wave fronts.

Fig. 2. Magnetic resonance imaging of spin waves above and below the superconducting transition temperature. (**A**) Spatial map of the NV photoluminescence PL_0 in the absence of microwaves, showing the MoRe strip (between the vertical dashes). Scale bar, 10 μm. (**B** and **C**) Spatial maps of the NV electron spin resonance contrast C_1 above (B) and below (C) the superconducting transition temperature of T_c = 8.7 K at T = 10.7 and 5.5 K, respectively. The Au, which excites spin waves, is located just outside the left edge of the imaged area. Above T_c, the wavelength is unaffected by the MoRe strip; below T_c, it is lengthened. (**D**) DC resistance R of the MoRe strip as a function of temperature T, with markers indicating the resistance of the film during the measurements of (B) triangle and (C) square. (**E**) Data from (B) and (C) averaged over the z direction, with the MoRe strip indicated by yellow shading. (**F**) Calculated spin-wave dispersion $f_{YIG}(k)$ for bare YIG, and $f_{YIG/SC}(k, \lambda_L)$ for YIG, covered by a superconducting film with London penetration depth λ_L = 400 nm. The superconductor shifts the dispersion upwards by $f_{SC}(k, \lambda_L)$, which manifests as a reduction in the wave number at NV frequency f_1 from $k^{(1)}_{YIG}$ to $k^{(1)}_{YIG/SC}$, indicated by the dashed lines.



By integrating this field self-consistently into the Landau-Lifshitz-Gilbert equation, we found that the spin-wave dispersion shifts upwards in frequency as

$$f_{\mathrm{YIG/SC}}(k,\lambda_{\mathrm{L}}) = f_{\mathrm{YIG}}(k) + f_{\mathrm{SC}}(k,\lambda_{\mathrm{L}}) \quad (1)$$

where $f_{\mathrm{YIG}}(k)$ is the bare-YIG spin-wave dispersion (SI) and

$$f_{\mathrm{SC}}(k,\lambda_{\mathrm{L}}) \approx \gamma\mu_0 M_s k t r \frac{1-e^{-2h/\lambda_{\mathrm{L}}}}{(k\lambda_{\mathrm{L}}+1)^2-(k\lambda_{\mathrm{L}}-1)^2 e^{-2h/\lambda_{\mathrm{L}}}} \quad (2)$$

is the superconductor-induced shift [supplementary text, section S1 (*19*)]. Here, M_s is the YIG saturation magnetization, t = 245 nm is the YIG thickness, h = 140 nm is the superconductor thickness, γ = 28 GHz/T is the electron gyromagnetic ratio, μ_0 is the vacuum permeability, and r is a dimensionless factor associated with the YIG thickness and spin-wave ellipticity. The approximation holds when the kinetic inductance dominates the impedance, as is the case for our superconducting strip [supplementary text, section S1 (*19*)], and when $k^2\lambda_{\mathrm{L}}^2 \ll 1$. A more general expression is given in the supplementary text (*19*). The dispersion shift $f_{\mathrm{SC}}(k,\lambda_{\mathrm{L}})$ is maximal when $\lambda_{\mathrm{L}} \to 0$, in which case the superconductor perfectly screens the spin-wave stray field. This limit was analyzed in (*25*, *26*) by considering a magnetic film covered by a "perfect metal" as defined by a perfect magnetic field screening. Indeed, when we let $\lambda_{\mathrm{L}} \to 0$, the shift calculated by our model approached the shift predicted in (*19*, *25*, *26*). The calculated bare YIG and hybridized YIG-MoRe spin-wave dispersions are compared in Fig. 2F. The upwards frequency shift underneath the superconductor manifested as a reduction in wave number of the NV-resonant spin waves detected in our experiments (Fig. 2E).

We observed that the ESR contrast just to the right of the MoRe strip was higher than that in the MoRe-strip region (Fig. 2C), consistent with the screening of the spin-wave field by the Meissner currents. In addition, we observed that the ESR contrast just to the right of the MoRe strip exceeded that just to the left of it (Fig. 2E), which indicated the excitation of additional, secondary spin waves by the MoRe strip itself. This enhanced excitation of secondary spin waves is presumably caused by an additional microwave current in the MoRe strip that is excited through the direct geometric inductive coupling with the gold microstrip when the MoRe impedance changes as it is cooled below T_c.

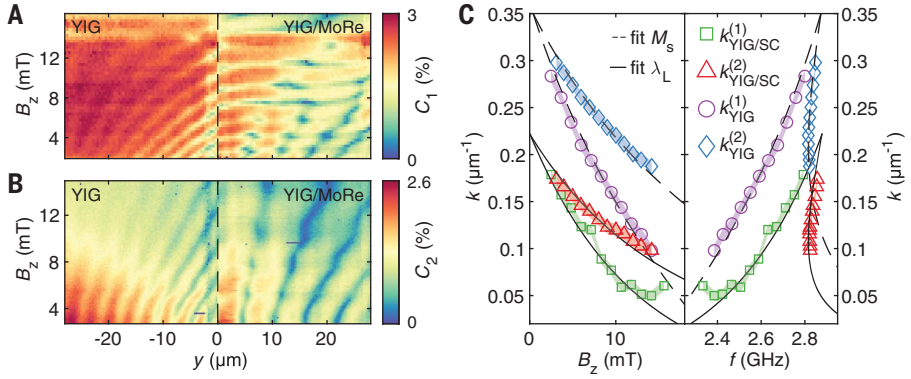


Fig. 3. Magnetic field dependence of the spin-wave dispersion in the magnet-superconductor hybrid. (**A** and **B**) Spatial line traces of the NV ESR contrasts C_1 and C_2 as a function of magnetic field B_z. Spin waves excited in the bare YIG region (y < 0) by the left Au microstrip (outside the imaged area) travel toward and then underneath the superconducting strip (y > 0), changing their wavelength. Interference with secondary spin waves excited at the MoRe strip edge (at y = 0) due to inductive coupling between Au and MoRe strips yields a beating pattern along B_z for y > 0. T = 5.5 K. The drive frequency is adjusted at each B_z to maintain resonance with the NV ESR transitions. (**C**) Spin-wave number as a function of field (left) and frequency (right), extracted from the data in (A) and (B) by means of Fourier transformation (fig. S2). The $k^{(i)}$ are the wave numbers measured with the field-aligned (i = 1) and misaligned (i = 2) NV ensembles. The error bars (indicated by shading) are determined by the inverse of the spatial sampling range in the y direction. We determined the saturation magnetization M_s by fitting the data in the bare YIG region, and the London penetration depth λ_L by fitting the data in the YIG-MoRe region using our YIG/SC model [supplementary text, section 2 (*19*)].

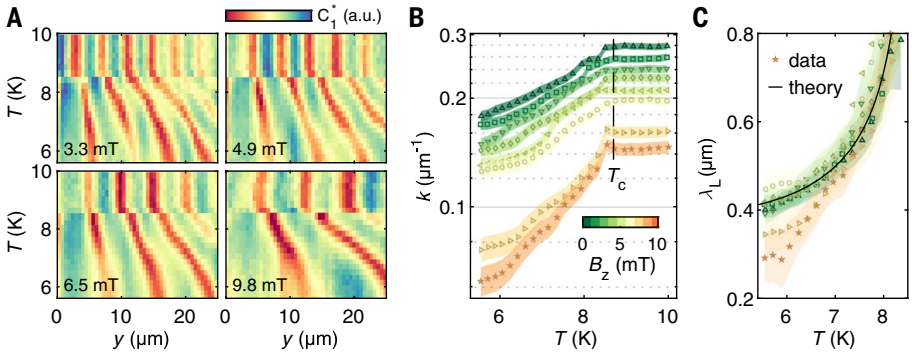


Fig. 4. Temperature tunability of the hybrid spin-wave–Meissner-current dispersion and extraction of the London penetration depth. (**A**) Spatial line traces of the NV ESR contrast C_1^* across the YIG-MoRe region as a function of temperature, showing the continuous change of the spin wavelength underneath the MoRe strip, for different in-plane magnetic fields B_z. The data are linearly detrended along y. Above the superconducting phase transition there is no temperature dependence of the wavelength, indicating the absence of the Meissner effect. A.u., arbitrary units. (**B**) Spin-wave numbers k extracted from data in (A) and from additional data in fig. S4 as a function of temperature. The colors indicate the different magnetic field values B_z. (**C**) London penetration depth λ_L of the MoRe film as a function of temperature, extracted from the data in (B) through our YIG/SC model. The black line represents the fit of the temperature dependence of $\lambda_L(T)$ from which we extracted T_c = 8.7 K and λ_L^0 = 380 nm. The colors indicate the different values of B_z as in (B).

Temperature- and field dependence of the spin-wave dispersion and extraction of the London penetration depth

We characterized the magnetic-field dependence of the spin-wave dispersion underneath the superconductor and used it to extract the London penetration depth λ_L at the T = 5.5 K base temperature of our cryostat. Spatial line traces of the NV ESR contrast across the strip shows the dependence of the spin wavelength on the applied magnetic field for the field-aligned (Fig. 3A) and misaligned (Fig. 3B) NV ensembles. In both measurements, we adjusted the drive frequency at each magnetic field to maintain resonance with the NV ESR frequency. We extracted the spin-wave numbers in the bare YIG and YIG-MoRe regions separately by Fourier transformation (fig. S2) and plotted these as a function of field and frequency in Fig. 3, C and D. A similar measurement with the bias field applied in plane along $\hat{\boldsymbol{x}}$ shows that Meissner screening of the bias field does not play a significant role in the wavelength shift (fig. S3).

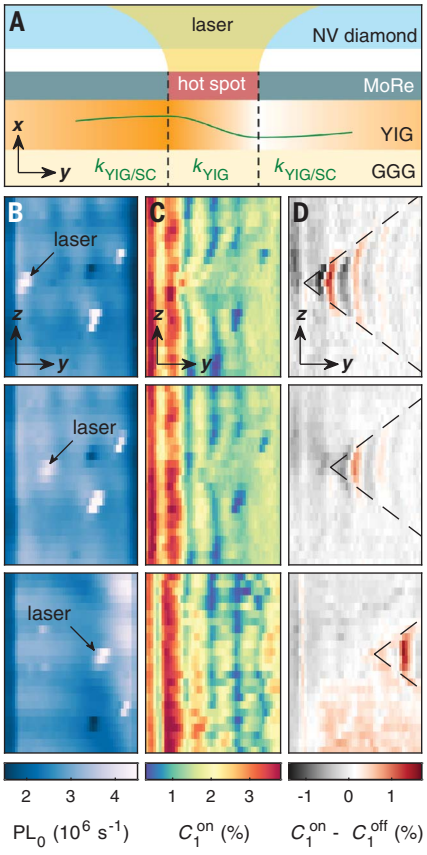


Fig. 5. Laser-induced spin-wave refraction at target locations. (**A**) Schematic illustration of a laser-induced scattering spot. By shining an auxiliary 594-nm laser on the sample, we created a hot spot in the MoRe strip that locally altered the effective refractive index governing the spin-wave propagation. (**B**) Scanning confocal microscope images of the NV photoluminescence at T = 5.5 K, with the auxiliary laser focused onto the MoRe strip at three different locations indicated by the arrows. Scale bar, 10 μm. (**C**) Spatial maps of the NV ESR contrast with auxiliary laser turned on (C_1^{on}) showing spin waves in the YIG-MoRe region that scatter on the laser spot. (**D**) Background-subtracted ESR contrast highlighting the laser-induced spin-wave scattering obtained by subtracting the ESR contrast with the auxiliary laser turned off C_1^{off} from the measurements in (C). B_z = 3.3 mT.

From the field dependence of the extracted spin-wave numbers in the bare YIG region, we extracted the YIG saturation magnetization M_s = 194(1) kA/m [supplementary text, section S2 (*19*)] in agreement with previous low-temperature measurements (*27*). We then used M_s as a fixed parameter to fit the field dependence of the spin-wave numbers underneath the superconducting strip using the hybridized YIG-MoRe spin-wave dispersion (Eq. 1). From this fit, we extracted a London penetration depth λ_L = 405(10) nm at T = 5.5 K, which agrees well with static-field nano–superconducting quantum interference device measurements (*28*).

The temperature dependence of the London penetration depth provides a powerful tool for tuning the spin wavelength. To demonstrate this, we imaged the spin waves in the YIG-MoRe region while sweeping through T_c at different magnetic fields (Fig. 4A and fig. S4). The extracted spin-wave number k is shown in Fig. 4B, with the color indicating the in-plane component of the magnetic field. We observed that k changes continuously with temperature over the superconducting phase transition in the YIG-MoRe region and remains unchanged in the bare YIG region (fig. S5). We did not observe global heating of the superconductor due to our excitation laser (fig. S6). Using our model, we extracted the London penetration depth $\lambda_L(T)$ for every observed value of k (Fig. 4C) [supplementary text, section S2 (*19*)]. We found that almost all data collapse onto a single curve described by $\lambda_L = \lambda_L^0[1-(T/T_c)^4]^{-1/2}$ (*29*), with T_c = 8.7 K and λ_L^0 = 380nm. The exceptions occurred when the spin wavelength $\lambda_{SW} = 2\pi/k$ became comparable to the width of the MoRe strip. Here, our approximation of the superconducting strip by an infinite film breaks down. These results highlight that imaging the hybridized spin-wave–Meissner-current transport modes is a powerful tool for extracting the temperature dependence of the London penetration depth.

Local control of spin-wave transport by laser-induced spin-wave refraction

Thus far, we have demonstrated dispersion engineering through global control of temperature and magnetic field. We now show that the creation of a hot spot in the superconductor with a focused laser enables local manipulation of the spin-wave transport by tuning the effective refractive index (Fig. 5A). To do so, we coupled an auxiliary, orange laser into our setup and focused it at target sites on the superconductor (Fig. 5B). The laser spot is visible through the locally enhanced NV photoluminescence. Spatial measurements of the NV ESR contrast C_1^{on} with the auxiliary laser on (Fig. 5C) show the spin-wave scattering patterns induced by the local hot spot. The reduction in amplitude behind the hot spot indicates destructive interference between the scattered and incident spin waves. Subtracting a reference measurement with the auxiliary laser turned off (Fig. 5D) highlights the angular profile of the scattered spin-wave patterns.

The characteristic "caustic" angles observed in these scattered patterns (dashed lines, Fig. 5D) result from the highly anisotropic dipolar spin-wave dispersion (*30*). Tracing the patterns to their origin shows that the scattering site is tightly confined to the laser location. Presumably, the laser locally breaks the superconductivity, inducing a local change in the magnetic environment seen by the spin waves, leading to local spin-wave refraction akin to defect-controlled spin-wave scattering (*30*). The ability to optically induce spin-wave refraction at target sites could be used to create devices such as gratings or magnonic crystals (*31*) and enable spin-wave manipulation through optical switching of flux-focusing regions in the superconducting strip.

Conclusions

We demonstrated local measurements of hybridized spin-wave–Meissner-current transport modes in a magnetic thin film equipped with a superconducting gate. The wavelength was tunable by temperature and field, enabling efficient phase-shifting of the spin-wavefronts and a notable in situ visualization and quantitative extraction of the London penetration depth as a function of temperature. Because MoRe is a type-2 superconductor with an estimated lower critical field of $H_{c1} = \frac{\Phi_0}{4\pi\lambda_L^2}\left[\ln\left(\frac{\lambda_L}{\xi}\right)+0.5\right] \approx$ 4 mT at $T \approx 5$ K (*32*, *33*), where $\xi \approx 0.01$ μm is the coherence length (*34*) and Φ_0 the superconducting flux quantum, Abrikosov vortices were expected in measurements such as those in Fig. 3B with a few-mT out-of-plane field component. However, we did not identify vortex-related effects in these measurements, which look qualitatively similar to those with purely in-plane field (fig. S3). Presumably, the presence and location of vortices is strongly influenced by our focused excitation laser, as highlighted by recent magneto-optical (*35*) and wide-field NV-imaging experiments (*36*). In particular, (*35*) showed that vortices can be annihilated by the laser through local heating above T_c or pushed by the laser to new pinning sites or border regions of the superconductor. The presented microwave magnetic imaging of the spin-wave transport modes in a YIG-MoRe heterostructure shows the versatility of superconducting gates for spin-wave manipulation, enables determining the temperature-dependent London penetration depth, and opens new opportunities for creating wave-based circuit elements such as filters, mirrors, and cavities.

REFERENCES AND NOTES

1. S. A. Wolf *et al.*, *Science* **294**, 1488–1495 (2001).
2. A. V. Chumak, V. I. Vasyuchka, A. A. Serga, B. Hillebrands, *Nat. Phys.* **11**, 453–461 (2015).
3. A. Barman *et al.*, *J. Phys. Condens. Matter* **33**, 413001 (2021).
4. J. Sinova, S. O. Valenzuela, J. Wunderlich, C. H. Back, T. Jungwirth, *Rev. Mod. Phys.* **87**, 1213–1260 (2015).
5. L. J. Cornelissen, J. Liu, R. A. Duine, J. Ben Youssef, B. J. Van Wees, *Nat. Phys.* **11**, 1022–1026 (2015).
6. A. Hamadeh *et al.*, *Phys. Rev. Lett.* **113**, 197203 (2014).
7. T. Yu, C. Liu, H. Yu, Y. M. Blanter, G. E. W. Bauer, *Phys. Rev. B* **99**, 134424 (2019).
8. J. Chen *et al.*, *Phys. Rev. Lett.* **120**, 217202 (2018).
9. Y. Sun *et al.*, *Phys. Rev. Lett.* **111**, 106601 (2013).
10. I. Bertelli *et al.*, *Adv. Quantum Technol.* **4**, 2100094 (2021).
11. Z. Y. Zhang *et al.*, *Supercond. Sci. Technol.* **25**, 025009 (2012).
12. A. Blais, A. L. Grimsmo, S. M. Girvin, A. Wallraff, *Rev. Mod. Phys.* **93**, 025005 (2021).

13. I. A. Golovchanskiy *et al*., *J. Appl. Phys.* **127**, 093903 (2020).
14. O. V. Dobrovolskiy *et al*., *Nat. Phys.* **15**, 477–482 (2019).
15. T. Yu, G. E. W. Bauer, *Phys. Rev. Lett.* **129**, 117201 (2022).
16. L. Rondin *et al*., *Rep. Prog. Phys.* **77**, 056503 (2014).
17. T. X. Zhou *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019473118 (2021).
18. I. Bertelli *et al*., *Sci. Adv.* **6**, eabd3556 (2020).
19. Materials and methods, and supplementary text, figures and references are available as supplementary materials.
20. F. Casola, T. Van Der Sar, A. Yacoby, *Nat. Rev. Mater.* **3**, 17088 (2018).
21. R. Schirhagl, K. Chang, M. Loretz, C. L. Degen, *Annu. Rev. Phys. Chem.* **65**, 83–105 (2014).
22. D. R. Glenn *et al*., *Geochem. Geophys. Geosyst.* **18**, 3254–3267 (2017).
23. P. Andrich *et al*., *npj Quantum Inf.* **3**, 28 (2017).
24. C. S. Wolfe *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 180406 (2014).
25. A. G. Gurevich, G. A. Melkov, *Magnetization Oscillations and Waves* (CRC Press, 1996).
26. S. R. Seshadri, *Proc. IEEE* **58**, 506–507 (1970).
27. S. Knauer *et al*., *J. Appl. Phys.* **133**, 143905 (2023).
28. A. G. Shishkin *et al*., *Supercond. Sci. Technol.* **33**, 065005 (2020).
29. M. Tinkham, *Introduction to superconductivity* (Dover Publications, 2nd edition, 2004).
30. R. Gieniusz *et al*., *Appl. Phys. Lett.* **102**, 102409 (2013).
31. A. V. Chumak, A. A. Serga, B. Hillebrands, *J. Phys. D Appl. Phys.* **50**, 244001 (2017).
32. W.-K. Kwok *et al*., *Rep. Prog. Phys.* **79**, 116501 (2016).
33. V. E. Calado *et al*., *Nat. Nanotechnol.* **10**, 761–764 (2015).
34. K. Makise *et al*., *Mater. Res. Express* **5**, 096406 (2018).
35. I. S. Veshchunov *et al*., *Nat. Commun.* **7**, 12801 (2016).
36. S. E. Lillie *et al*., *Nano Lett.* **20**, 1855–1861 (2020).
37. M. Borst *et al*., Data for: Observation and control of hybrid spin-wave–Meissner-current transport modes, version 1, Zenodo (2023).

ACKNOWLEDGMENTS

We thank A. F. Otte for commenting on the manuscript and T. Yu and G.E.W. Bauer for discussions. **Funding:** This work was supported by the Dutch Research Council (NWO) under awards VI.Vidi.193.077, NGF.1582.22.018, and OCENW.XL21.XL21.058. This work was also supported by the Dutch National Growth Fund (NGF) as part of the Quantum Delta NL programme. **Author contributions:** Conceptualization: T.S. and M.B.; Sample design and fabrication: M.B.; Diamond membrane fabrication: B.S. and M.B.; Membrane placement: A.L. and S.K.; Constructing experimental setup: M.B. and A.L.; Measurements and software: M.B., P.H.V., A.L., and I.B.; Data analysis: M.B., P.H.V., and A.T.; Theory: T.S. and Y.M.B.; Figures: M.B. and P.H.V.; Funding acquisition: T.S.; Project administration: M.B. and T.S.; Supervision: T.S. and M.B.; Writing – original draft: M.B., T.S., and P.H.V.; Writing – review and editing: M.B., T.S., P.H.V., and Y.M.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data presented in this work are available at Zenodo (*37*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj7576
Materials and Methods
Supplementary Text
Figs. S1 to S6
References (*38*–*42*)

Submitted 14 July 2023; accepted 18 September 2023
10.1126/science.adj7576

PHOTONICS

Quantum walk comb in a fast gain laser

Ina Heckelmann†*, Mathieu Bertrand†, Alexander Dikopoltsev†, Mattias Beck, Giacomo Scalari, Jérôme Faist*

Synthetic lattices in photonics enable the exploration of light states in new dimensions, transcending phenomena common only to physical space. We propose and demonstrate a quantum walk comb in synthetic frequency space formed by externally modulating a ring-shaped semiconductor laser with ultrafast recovery times. The initially ballistic quantum walk does not dissipate into low supermode states of the synthetic lattice; instead, the state stabilizes in a broad frequency comb, unlocking the full potential of the synthetic frequency lattice. Our device produces a low-noise, nearly flat broadband comb (reaching 100 per centimeter bandwidth) and offers a promising platform to generate broadband, tunable, and stable frequency combs.

The random walk is a fundamental concept in mathematics and physics that describes a walker taking random steps in a discretized space. The spread of a classical random walker is characterized by a $\sqrt{N}$ standard deviation after N steps (*1*). It can, however, be surpassed by its quantum counterpart, which permits the particle to behave as a wave and interfere with itself (*2*). In this quantum walk the standard deviation of the position scales linearly with the number of steps, enabling a so-called ballistic spread that facilitates the "quantum speed-up" of search algorithms based on quantum walks (*3*). As coherent light mimics the interference of quantum particles while benefiting from the simpler preparation and manipulation of the light states, photonic lattices provide a natural platform to study quantum walks (*4*, *5*). This line of research has produced phenomena such as quantum-to-classical walk transitions (*6*) and quantum walks of correlated photons (*7*), as well as the experimental study of boson interactions in photonic systems (*8*).

In such systems the study of quantum walks of light was limited to the type of interactions and nonlinearities that experiments in real space could provide. However, by coupling accessible degrees of freedom of light, it is possible to construct a synthetic dimension that enables the realization of previously unattainable physical phenomena. This approach leverages the flexibility and controllability of synthetic dimensions to engineer photonic lattices that effectively overcome the geometric limitations of the system (*9*–*12*). In this way, synthetic dimensions allow for the observation of physics such as parity-time-symmetric lattices (*13*), complex-valued long-range couplings (*14*), higher-dimensional photonic crystals (*15*, *16*), non-Hermitian funneling of light (*17*), and even three-dimensional (3D) topological insulators (*18*).

For a synthetic dimension formed by the modes of a resonator, such as frequency ladders in optical ring cavities (Fig. 1A), a phase modulation (induced, for instance, through the electro-optical effect or the laser's pump) can introduce coupling in the chain, giving rise to continuous-time quantum walk dynamics (*5*). For a fixed coupling C, the broadest achievable state is limited by the dispersion in the system, which causes mode resonances to n apart, resulting in less efficient coupling (1B) (*11*). Qualitatively, dispersion induces an effective potential well in frequency space whereas the width of the band created by modulation limits the maximum achievable "kinetic energy" and, consequently, the system's bandwidth. Through this argument, for instance, quadratic dispersion will constrain the highest accessible mode to $n_{max} = \sqrt{2\frac{M}{D}}$, where $M = 2C$ represents the modulation strength and D signifies dispersion. This system can be analogously mapped to a quantum harmonic oscillator, which could only occupy a finite number of possible modes (Fig. 1C).

In harmonic oscillator solutions only the 0th order mode is a minimum uncertainty state, being narrowest both in frequency space and reciprocal time whereas modes higher on the energy ladder exhibit simultaneously broader spectral and temporal widths (Fig. 1, C and D). Introducing linear non-Hermiticity to the system in the form of frequency-dependent gain and dissipation leads to a contraction of the final state to a narrow Gaussian profile (i.e., a pulse) when reaching the state at the bottom of the dispersion trap. Dynamically, this behavior is characterized by a transition of the initial ballistic broadening of the quantum walk to a regime of diffusive propagation (*6*, *19*) (Fig. 1E). Previously, a similar mapping considering gain curvature instead of dispersion was established (*20*), demonstrating the dissipation to low supermodes in frequency lattices and an $M^{1/4}$-dependence of the bandwidth (*21*) (fig. S13). The stabilization in low modes ex-

Institute of Quantum Electronics, ETH Zürich, 8093 Zürich, Switzerland, and Quantum Center, ETH Zürich, 8093 Zürich, Switzerland.
*Corresponding author. Email: iheckelmann@ethz.ch (I.H.); jfaist@ethz.ch (J.F.)
†These authors contributed equally to this work.

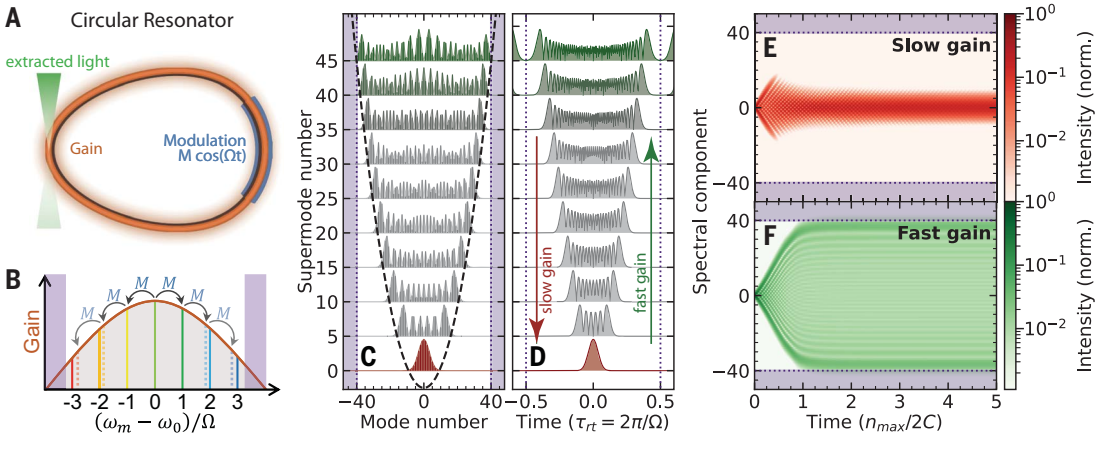


Fig. 1. Stabilization of quantum walks in synthetic photonic lattices. The optical modes in a deformed circular cavity (**A**) are coupled through resonant modulation, inducing nearest-neighbor gain and the proliferation of a synthetic photonic lattice in frequency space, whose bandwidth is fundamentally limited by the dispersion (indigo) (**B**). In analogy to the quantum harmonic oscillator, the 0th order supermode corresponds to a Gaussian pulse (red) in the discretized frequency (**C**) and continuous time domain (**D**), whereas constant intensity can only be seen for spectrally broad supermodes (green). After the initial ballistic expansion upon the onset of nearest-neighbor coupling *C*, dissipation gives rise to a stabilization of the spectrum. Although it collapses to a narrow Gaussian mode for typical slow-gain dissipative systems (**E**), a superposition of broad supermodes can be observed when a fast gain is employed instead (**F**).

poses the inherent limitations of such systems in reaching their intrinsic bandwidth.

Generating the synthetic frequency lattice

We demonstrate both theoretically and experimentally that the full potential of the synthetic frequency lattice can be unlocked by employing ultrafast saturable gain. The giant third-order nonlinearity of such gain locks the resonator modes and effectively counteracts the dispersion. After the initial single mode performs a spectral quantum walk and expands ballistically, the laser spectrum stabilizes on the highest energy supermodes, reaching the predicted fundamental bandwidth limit with a flat spectral envelope and near-constant intensity operation (Fig. 1F). This is in contrast to broadband optical frequency combs (*22*–*24*) based on pulsed mode-locked lasers (*25*) or on nonlinear generation in microresonators (*26*, *27*), which typically stabilize into a peaked spectrum characteristic of a pulsed state. Additionally, in the proposed quantum walk comb, the bandwidth of the resulting low-noise frequency comb is shown to be continuously tunable by varying the radio frequency (RF) injection.

We consider a synthetic lattice comprised of the resonator modes of an active, slightly deformed circular cavity (Fig. 1A and figs. S1 and S2), realized using a quantum cascade laser (QCL) emitting in the mid-infrared spectral region. QCLs are semiconductor lasers that provide a flexible platform for high-power combs in the molecular fingerprint region with small device footprints (*28*). Notably, their gain mechanism is based on intraband transitions and thus differs fundamentally from typical interband lasers, resulting in extremely fast gain recovery times in the order of hundreds of femtoseconds. This type of gain in a Fabry-Perot device will yield frequency modulated (FM) combs (*29*), created by four-wave mixing nonlinearities arising from spatial hole burning (*30*, *31*). To enable linear coupling among lattice sites in a synthetic space and observe the continuous-time quantum walk, we adopted a circular cavity geometry that suppresses spatial hole burning. Circular QCLs have recently attracted considerable interest since they were shown to host the formation of self-emerging cavity solitons—previously seen in passive microresonators (*32*)—resulting from the delicate balance between anomalous group velocity dispersion and gain nonlinearity (*33*–*36*). For our work, the lasers were fabricated using an inverted buried-heterostructure process to achieve smooth sidewalls and hence guarantee unidirectional single-mode lasing (*37*). In circular cavities where backscattering and hence coupling between counter-propagating modes is sufficiently low, spontaneous symmetry breaking (*33*) occurs close to above the threshold, enabling the laser to operate in a unidirectional single-mode state (fig. S3) (*38*). As a result, it is only when we introduce additional RF modulation at the cavity resonance frequency that a frequency comb with a broad, highly tuneable, and predictable bandwidth emerges.

When light is extracted from such circular cavities using only bending losses, the output is isotropic and was previously shown to be restricted to the milliwatt level (*33*). To vary the cavity circumference and consequently the roundtrip time without experiencing an exponential decrease in outcoupled light intensity, we adapted the geometry of a Hügelschäffer egg (table S1 and fig. S2) (*39*). In doing so, the higher curvature at the tip of the eggs creates a designated out-coupling point. Coupling of the resonator modes in the ring QCL was achieved through near-resonant RF current modulation, which provides an efficient phase modulation of the lasing electric field. This effect is enabled by the active region itself, as gain modulation is translated into phase modulation through a nonvanishing linewidth enhancement factor (*40*, *41*).

Characterizing the spectral output

To study quantum walk combs, steady-state spectra of our device were measured with Fourier-transform infrared spectroscopy (FTIR) for both a sweep of the injected frequency through the cavity resonance at 15.772 GHz, and without any applied modulation. Although the free-running device operates as a single-mode laser at 1280 cm^{-1} at the investigated working point at 1.456 A and −16°C (fig. S3C), it gradually and reliably broadens to a bandwidth of 70 cm^{-1} when RF modulation is enabled and tuned to resonance (Fig. 2B). For resonant injection (working point 2), the spectral envelope showcases lobes on both sides and has a nearly flat center, as expected for a quantum walk stabilizing on a highly excited state (*1*, *2*). For this resonant injection, a participation ratio given by $\frac{(\Sigma_m I_m)^2}{\Sigma_m I_m^2}$ where I_m is the intensity of the m-th optical mode of 74.3 is computed for the 132 involved lattice sites, indicating a highly delocalized (i.e., flat) spectrum. As we detune from resonance, the spectra narrow continuously and display an imbalance (working points 3 and 4), which rectifies far from resonance (working point 1).

To theoretically investigate the comb formation, we modeled the propagating field in the dispersive waveguide of the active ring QCL based on (*42*, *43*), while neglecting spatial hole burning due to the absence of any counter-propagating wave. Here, the phase of the field follows the modulation with depth M and detuning $\Delta\Omega$ from the cavity resonance. Through this model, assuming instantaneous gain for QCLs and thus a constant intensity, we calculated the steady state of the electric field in the corotating frame of reference [derivation in (*44*)]. By adjusting only a small set of parameters, we accurately reproduced the experimental spectra (Fig. 2, A and C). The bandwidth of the comb under resonant injection reaches the previously discussed fundamental limit of the synthetic lattice.

Upon closer examination of the phases predicted by our model, we observe the existence of two distinct lasing regimes, which depend on the detuning $\Delta\Omega$ of the injected frequency from the cavity resonance (fig. S11). In the case of highly off-resonant injection, the comb state exhibits Bessel-like spectra and sinusoidal instantaneous frequency with the periodicity of the injected signal, akin to the characteristics previously observed in pure FM lasers (*45*). However, within a range of near-resonant injections, we discovered a novel FM regime characterized by a discontinuous cosine phase at only half the modulation frequency (Fig. 2D). The phase adheres to the following analytically derived relation ("ana" in Fig. 2D), assuming a constant intensity:

$$\Phi(z,t) = N_1 \cdot \left(\frac{Kz - \Delta\Omega t}{2}\right) + N_2 \cdot \cos\left(\frac{Kz - \Delta\Omega t}{2}\right) \quad (1)$$

where K represents the cavity wave number, and N_1 and N_2 denote the amplitudes of phase modulation, which directly rely on the modulation depth M and dispersion β [derivation in (*44*)].

To experimentally corroborate the presence of these separate lasing regimes, we conducted phase measurements of the laser using shifted-wave interference Fourier-transform spectroscopy (SWIFTS) (*46*) at various injection frequencies around the resonance (exp in Fig. 2D and fig. S4). The measured instantaneous frequencies align well with the numerically simulated and analytically derived solutions for both lasing regimes. Furthermore, coherence was confirmed through these measurements, encompassing the entire spectral bandwidth (fig. S6). However, it is important to note that our experiment measures the collective behavior of the modes, resembling that of a single particle, and does not detect the presence of nonclassical correlations. Our steady-state measurements of the frequency comb convincingly confirmed its coherence, continuous and predictable tuning, wide bandwidth, and excellent agreement with our theoretical models.

Resolving the photonic quantum walk

Investigating the role of quantum walk dynamics in shaping the steady-state spectra, we conducted measurements and simulations to explore the underlying transient processes. To track the dynamical comb broadening under RF injection, we employed boxcar averaging in conjunction with an FTIR at 10 ns temporal

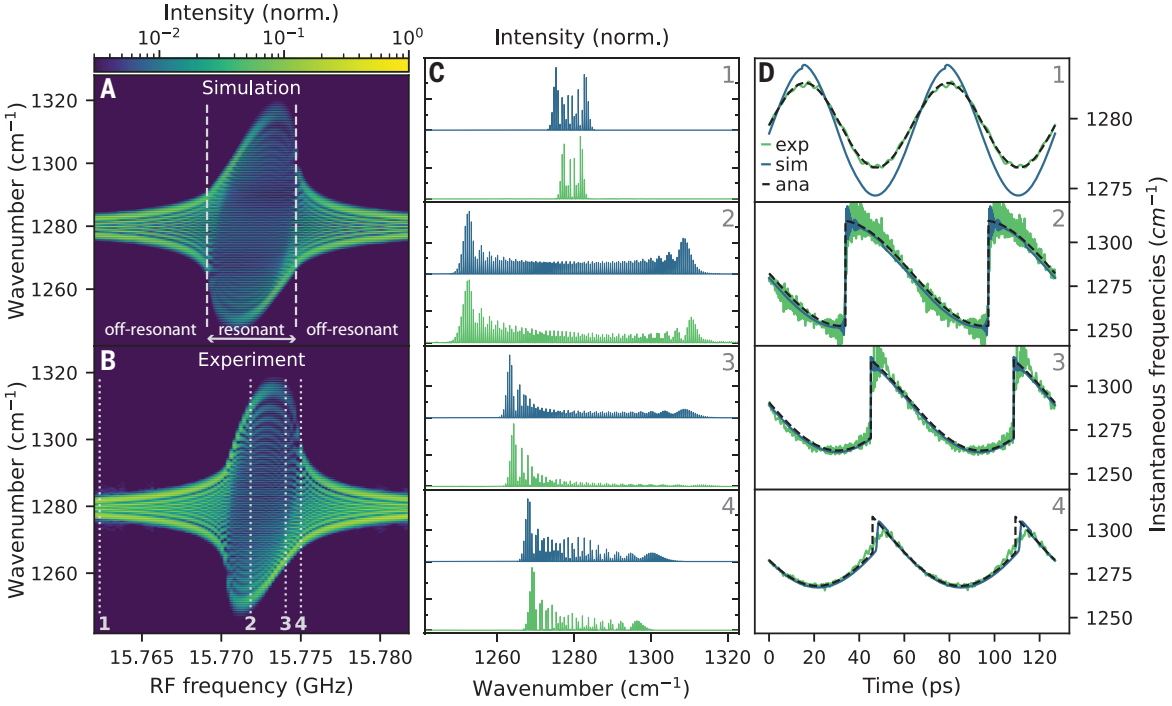


Fig. 2. Steady state of the quantum walk comb. Spectra for sweeps of the injection frequency, as predicted by simulations based on the Maxwell-Bloch equations (**A**) and measured in FTIR (**B**) with distinct on- and off-resonant lasing regimes. Spectra (**C**) (linear scale) and instantaneous frequencies (**D**) at the indicated working points [vertical lines in (B)] display narrow Bessel-like spectra and sinusoidal phases in the off-resonant regime [panel 1 in (D)] and discontinuous phases at half the roundtrip frequency in the novel on-resonance regime (panels 2 and 3). The transition between the two lasing states occurs at a detuning of 2.8 MHz (panel 4) with unbalanced spectra and an instantaneous frequency with negligible jumps. The experimental instantaneous frequencies (exp, green) are well-described by both numerical simulations (sim, blue) and the derived analytical expression (ana, black).

resolution (fig. S7 and S8). The experimental spectra (Fig. 3, C and D) exhibit a ballistic expansion to the full 70 cm^{-1} bandwidth within the initial 250 ns, followed by a subsequent locking that maintains this spectral state for the investigated times.

Building upon these experimental findings, we now delve into the underlying theoretical model using the Maxwell-Bloch framework to unravel the dynamics of the quantum walk responsible for the observed spectral broadening during RF modulation (Fig. 3, A and B). As mentioned earlier, the free-running laser solely occupies the resonator mode with the highest gain but when resonant injection is introduced, mixing of the optical photons and modulation quanta Ω leads to nearest-neighbor gain in the photonic lattice. Through discrete hopping between resonator modes, the photons perform a continuous-time quantum walk, resulting in a ballistic broadening of the spectrum whose bandwidth increases linearly with the number of quantum walk steps. Once the spectrum reaches the frequency boundary dictated by the dispersion and modulation parameters, it stabilizes into a broad supermode both spectrally and temporally, maintaining this state for subsequent time intervals. The only clear difference between the experiment and our model is the much slower experimental decay of the central mode at 1287 cm^{-1} in the measured spectra. The otherwise notable agreement between our simulated and experimental

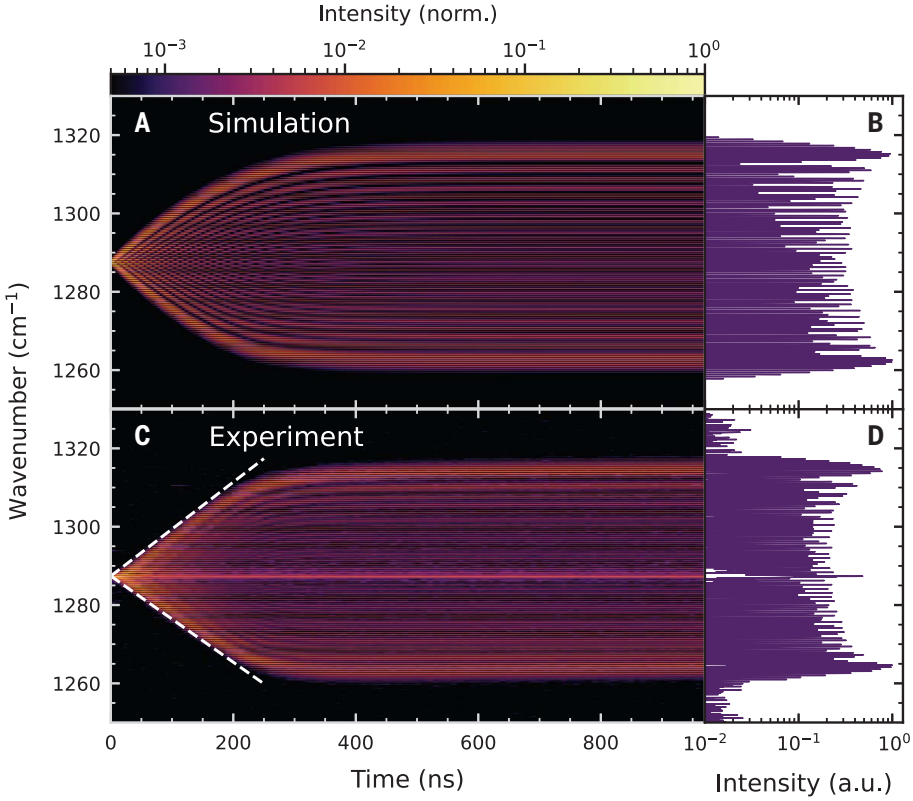


Fig. 3. Quantum walk and locking. Simulations based on the Maxwell-Bloch master equation (**A**) reproduce the ballistic expansion and subsequent locking to a spectrally broad supermode (**B**) that is observed in the time-resolved spectral measurements [(**C**) and (**D**)]. The dashed white lines indicate the ballistic broadening of the Quantum Walk.

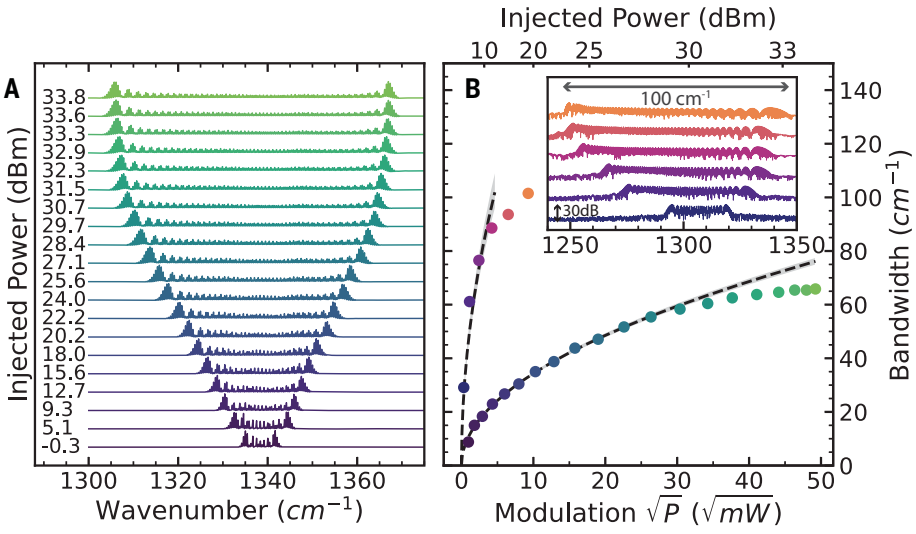


Fig. 4. Quantum walk comb device under RF injection. The progressive broadening of the spectra with increasing injection power [(**A**) (linear scale)] follows the predicted $M^{1/2}$ (and $P^{1/4}$) dependence up until saturation around 31.5 dBm (**B**), lower curve. A second device showed broadening up to 100 cm^{-1} but could not be driven at resonance, as shown by the typical asymmetric spectra (inset, logarithmic scale).

spectra provides compelling evidence for the rapid expansion and stable locking of the quantum walk comb.

Tunability and stability

Solving the Maxwell-Bloch equations at resonance, we can determine the state in which the quantum walk stabilizes, leading to a bandwidth dependence of $\sqrt{(M/D)}$ [derivation in (*44*)]. This power dependence of the bandwidth offers a quadratic improvement over the $M^{1/4}$-relation typically seen in active mode-locking of lasers (*21*). Because the modulation depth M is proportional to the square root of the injected power, a parameter that we experimentally control, we anticipate observing $P^{1/4}$-dependence of the bandwidth. Confirming this expectation, a power sweep at resonance validated the predicted scaling behavior, with saturation occurring around 66 cm^{-1} at 31.5 dBm for the investigated device (Fig. 4). We attribute the deviations from the expected behavior at high power to nonlinearities in the modulation dynamics of the QCL active region, as well as higher-order dispersion. The simple relationship between modulation power and optical bandwidth allows the power per comb line to be controlled flexibly and with very short time delays, because the average laser power remains approximately constant. On the same figure, we report the results obtained on a device which exhibited an even broader bandwidth that reached 100 cm^{-1}; however, it could not be driven to full resonance. The most likely reason is a larger amount of third-order dispersion resulting from the wider waveguide.

Under resonant injection, the comb demonstrates the same level of stability as the free-running single-mode laser, providing evidence that the modulation itself does not introduce additional noise contributions. In particular, the measured amplitude noise power spectral density is unchanged by modulation, is shot-noise limited above ~100 kHz, and exhibits the distinct 1/f-dependence for lower frequencies (fig. S9). These notable characteristics attest to the exceptional stability of the comb operation. With the inherent stability of the RF modulation faithfully transferred to the optical spectrum, the quantum walk comb presents a highly promising platform for the generation of broadband, tunable, and stable frequency combs.

Concluding remarks

We have observed in real time the ballistic expansion and stabilization of the spectrum in a synthetic photonic lattice operating in a laser upon resonant phase modulation, a display that can be ascribed to nonlinear continuous-time quantum walks. In particular we demonstrate that in our ring QCL structure—the fast gain saturation that constrains the output to a near-constant intensity—leads the quantum walk to the spectrally broadest state allowed by dispersion. The quantum walk comb exhibits high predictability, tunability, and stability. In this way, we address an issue raised many years ago: active mode-locked devices with slow gain recovery operate on the fundamental narrowest Gaussian mode with limited comb bandwidth, all other modes being unstable (*20*). Leveraging the insights gained from our work, deliberate design strategies for wide-bandwidth comb sources could open up applications in fields such as spectroscopy (*47*) and telecommunications (*48*), as the principles of this work are not limited to QCLs but could equally exploit fast gain components existing in interband devices to expand to other wavelength regimes (*49*). Furthermore, this approach holds promise for the application of quantum walk algorithms in systems featuring designated nonlinear interactions among qubits and can be generalized to higher dimensional lattices in either real space with coupled resonators or synthetic space using additional modulation frequencies (*16*).

REFERENCES AND NOTES

1. F. Spitzer, *Principles of Random Walk* (Springer, 2013).
2. Y. Aharonov, L. Davidovich, N. Zagury, *Phys. Rev. A* **48**, 1687–1690 (1993).
3. N. Shenvi, J. Kempe, K. B. Whaley, *Phys. Rev. A* **67**, 052307 (2003).
4. A. Schreiber *et al.*, *Phys. Rev. Lett.* **104**, 050502 (2010).
5. A. Rai, G. S. Agarwal, J. H. H. Perk, *Phys. Rev. A* **78**, 042304 (2008).
6. T. Eichelkraut *et al.*, *Nat. Commun.* **4**, 2533 (2013).
7. A. Peruzzo *et al.*, *Science* **329**, 1500–1503 (2010).
8. Y. Lahini *et al.*, *Phys. Rev. A* **86**, 011603 (2012).
9. A. Celi *et al.*, *Phys. Rev. Lett.* **112**, 043001 (2014).
10. T. Ozawa, H. M. Price, N. Goldman, O. Zilberberg, I. Carusotto, *Phys. Rev. A* **93**, 043827 (2016).
11. L. Yuan, Y. Shi, S. Fan, *Opt. Lett.* **41**, 741–744 (2016).
12. M. Ehrhardt, S. Weidemann, L. J. Maczewsky, M. Heinrich, A. Szameit, *Laser Photonics Rev.* **17**, 2200518 (2023).
13. A. Regensburger *et al.*, *Nature* **488**, 167–171 (2012).
14. B. A. Bell *et al.*, *Optica* **4**, 1433 (2017).
15. A. Dutt *et al.*, *Science* **367**, 59–64 (2020).
16. Y. Hu, C. Reimer, A. Shams-Ansari, M. Zhang, M. Loncar, *Optica* **7**, 1189 (2020).
17. S. Weidemann *et al.*, *Science* **368**, 311–314 (2020).
18. E. Lustig *et al.*, *Nature* **609**, 931–935 (2022).
19. A. Dikopoltsev *et al.*, *Sci. Adv.* **8**, eabn7769 (2022).
20. H. Haus, *IEEE J. Quantum Electron.* **11**, 323–330 (1975).
21. D. Kuizenga, A. Siegman, *IEEE J. Quantum Electron.* **6**, 694–708 (1970).
22. T. Udem, R. Holzwarth, T. W. Hänsch, *Nature* **416**, 233–237 (2002).
23. S. A. Diddams, K. Vahala, T. Udem, *Science* **369**, eaay3676 (2020).
24. L. Chang, S. Liu, J. E. Bowers, *Nat. Photonics* **16**, 95–108 (2022).
25. U. Keller, *Nature* **424**, 831–838 (2003).
26. P. Del'Haye *et al.*, *Nature* **450**, 1214–1217 (2007).
27. B. Stern, X. Ji, Y. Okawachi, A. L. Gaeta, M. Lipson, *Nature* **562**, 401–405 (2018).
28. J. Faist *et al.*, *Science* **264**, 553–556 (1994).
29. A. Hugi, G. Villares, S. Blaser, H. C. Liu, J. Faist, *Nature* **492**, 229–233 (2012).
30. J. B. Khurgin, Y. Dikmelik, A. Hugi, J. Faist, *Appl. Phys. Lett.* **104**, 081118 (2014).
31. A. Gordon *et al.*, *Phys. Rev. A* **77**, 053804 (2008).
32. T. J. Kippenberg, A. L. Gaeta, M. Lipson, M. L. Gorodetsky, *Science* **361**, eaan8083 (2018).
33. B. Meng *et al.*, *Nat. Photonics* **16**, 142–147 (2022).
34. S. Kacmoli, D. L. Sivco, C. F. Gmachl, *Opt. Express* **30**, 47475–47484 (2022).
35. M. Piccardo *et al.*, *Nature* **582**, 360–364 (2020).
36. P. Micheletti *et al.*, *Sci. Adv.* **9**, eadf9426 (2023).
37. M. Beck *et al.*, *Science* **295**, 301–305 (2002).
38. L. Seitner *et al.*, Backscattering-Induced Kerr Solitons in Ring Quantum Cascade Lasers. arXiv.2304.07054 [physics.optics] (2023).
39. H. Schmidbauer, *Elemente der Mathematik* **III**, 67 (1948).
40. N. Opačak, S. D. Cin, J. Hillbrand, B. Schwarz, *Phys. Rev. Lett.* **127**, 093902 (2021).
41. M. Franckié, M. Bertrand, J. Faist, *Appl. Phys. Lett.* **122**, 021107 (2023).
42. D. Burghoff, *Optica* **7**, 1781 (2020).
43. N. Opačak, B. Schwarz, *Phys. Rev. Lett.* **123**, 243902 (2019).
44. Materials and methods are available as supplementary materials.
45. S. Harris, O. McDuff, *IEEE J. Quantum Electron.* **1**, 245–262 (1965).

46. D. Burghoff *et al*., *Opt. Express* **23**, 1190–1202 (2015).
47. N. Picqué, T. W. Hänsch, *Nat. Photonics* **13**, 146–157 (2019).
48. J. Pfeifle *et al*., *Nat. Photonics* **8**, 375–380 (2014).
49. U. Senica *et al*., Frequency-modulated combs via on-chip field enhancement. arXiv.2305.01483 [physics.optics] (2021).
50. I. Heckelmann, M. Bertrand, A. Dikopoltsev, Quantum Walk Comb in a fast gain laser, ETH Research Collection (2023); https://doi.org/10.3929/ethz-b-000618126.

ACKNOWLEDGMENTS

Funding: This work was supported by the following: MIRAQLS: Staatssekretariat für Bildung, Forschung und Innovation SBFI (22.00182) in collaboration with EU (grant Agreement 101070700); Swiss National Science Foundation (212735); Innosuisse: Innovation Project 52899.1 IP-ENG (Agreement Number 2155008433 "High yield QCL Combs"); ETH Fellowship program: (22-1 FEL-46) (to A.D.)
Author contributions: I.H. processed the devices, performed the characterizations, and wrote the original draft. M.Ber. performed and analyzed the SWIFTS and time-resolved measurements with assistance from I.H. A.D. provided predictive models and performed the simulations. I.H., M.Ber., A.D., and M.Bec. wrote the Supplementary Materials. M.Bec. grew the QCL wafer and consulted in the fabrication. A.D., G.S., and J.F. acquired the funding, administrated, and supervised the project. J.F. conceptualized the idea. All authors contributed to the interpretation of the results and the review and editing of the draft.
Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** Data that support the findings of this article and codes are available in the ETH Research Collection (*50*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj3858
Materials and Methods
Supplementary Text
Figs. S1 to S13
Table S1

Submitted 23 June 2023; accepted 7 September 2023
10.1126/science.adj3858

SEMICONDUCTORS

Room-temperature wavelike exciton transport in a van der Waals superatomic semiconductor

Jakhangirkhodja A. Tulyagankhodjaev, Petra Shih†, Jessica Yu†, Jake C. Russell, Daniel G. Chica, Michelle E. Reynoso, Haowen Su, Athena C. Stenor, Xavier Roy, Timothy C. Berkelbach, Milan Delor*

The transport of energy and information in semiconductors is limited by scattering between electronic carriers and lattice phonons, resulting in diffusive and lossy transport that curtails all semiconductor technologies. Using $Re_6Se_8Cl_2$, a van der Waals (vdW) superatomic semiconductor, we demonstrate the formation of acoustic exciton-polarons, an electronic quasiparticle shielded from phonon scattering. We directly imaged polaron transport in $Re_6Se_8Cl_2$ at room temperature, revealing quasi-ballistic, wavelike propagation sustained for a nanosecond and several micrometers. Shielded polaron transport leads to electronic energy propagation lengths orders of magnitude greater than in other vdW semiconductors, exceeding even silicon over a nanosecond. We propose that, counterintuitively, quasi-flat electronic bands and strong exciton–acoustic phonon coupling are together responsible for the transport properties of $Re_6Se_8Cl_2$, establishing a path to ballistic room-temperature semiconductors.

Semiconductor technologies rely on transporting energy and information carriers, often in the form of electrons or excitons (bound electron-hole pairs), from source to target. At room temperature, these carriers rapidly scatter with lattice vibrations (phonons) on nanometer and femtosecond scales. Scattering leads to electronic energy dissipation, joule heating, and loss of phase coherence and directionality, imposing strict speed and efficiency limits on all semiconductor technologies. Breaking through these limits requires semiconductors that sustain ballistic (scatter-free), wavelike flow of energy over macroscopic distances at room temperature, a long-sought goal that would enable ballistic transistors (*1*), low-loss energy harvesting, and wave-based information technologies (*2*).

Here, we demonstrate macroscopic, wavelike exciton flow at room temperature in the van der Waals (vdW) superatomic material $Re_6Se_8Cl_2$ (Fig. 1A). $Re_6Se_8Cl_2$ is a semiconductor with an indirect bandgap of 1.6 eV and an exciton binding energy of ~100 meV (*3*). Its superatom building blocks consist of Re_6 octahedra enclosed in Se_8 cubes. Each Re_6Se_8 unit is covalently bonded to four neighbors to form a two-dimensional (2D) pseudosquare lattice capped by Cl atoms at the apical positions. The $Re_6Se_8Cl_2$ layers stack out of plane to create a bulk vdW crystal with weak interlayer electronic coupling (*4*). The crystal can be exfoliated to the monolayer limit, which is advantageous for integration in gated devices (*3*, *5*, *6*). $Re_6Se_8Cl_2$ exhibits relatively weak intercluster electronic coupling, as evidenced by the electronic band structure (Fig. 1B) (*3*, *7*). Strong coupling of electrons to intercluster optical phonons (*8*) leads to further band flattening at room temperature (*7*) and has been implicated in the emergence of superconductivity in this material and related classes (*6*, *9*). In this work, we demonstrate that these quasi-flat electronic bands, in combination with strong coupling to acoustic phonons, lead to the formation of acoustic exciton-polarons (Fig. 1C), quasiparticles of excitons bound to an acoustic lattice deformation. Through direct imaging of polaron propagation, we reveal that they are shielded from lattice scattering, leading to quasi-ballistic transport over several micrometers at room temperature, currently limited only by crystal size. Our observations challenge the common notion that strong electronic coupling is required for long-range transport.

Department of Chemistry, Columbia University, New York, NY 10027, USA.
*Corresponding author. Email: milan.delor@columbia.edu
†These authors contributed equally to this work.

Exciton transport in $Re_6Se_8Cl_2$ is quasi-ballistic

We directly imaged exciton transport in single-crystal $Re_6Se_8Cl_2$ using ultrafast stroboscopic scattering microscopy (stroboSCAT) (*10*–*12*) (Fig. 1D and figs. S1 to S3). An above-gap, diffraction-limited visible pump generates excitons, and then a backscattering widefield probe (1.55 eV) slightly below the electronic bandgap spatially resolves how the excitons modify the local polarizability of the material. By varying the pump-probe time delay, we spatiotemporally tracked the evolution of photoexcitations in an all-optical, noninvasive, and contact-free measurement. Figure 1E and movie S1 display representative stroboSCAT data obtained in a 60-nm-thick $Re_6Se_8Cl_2$ flake prepared by mechanical exfoliation (*13*). Two key features emerge from these data: First, the initial negative (dark) stroboSCAT contrast turns to positive (bright) contrast on picosecond timescales, which, as discussed below (Fig. 2), represents a transition from a bare exciton to an exciton-polaron. Second, the exciton-polaron propagates several micrometers to the edge of the flake in less than a nanosecond. This very fast and long-range transport differs sta… from exciton transport in other molecular o… semiconductors (table S1). For comparison, Fig. 1F displays stroboSCAT data of exciton transport in the archetypal vdW semiconductor WSe_2 (bilayer flake on glass; see fig. S4 for monolayer and bulk data), exhibiting much slower and shorter-range transport. These results are counterintuitive, given that the effective mass of excitons in WSe_2 is much smaller than in $Re_6Se_8Cl_2$ (table S1).

To quantify and rationalize the exceptional transport properties of $Re_6Se_8Cl_2$, we plotted the mean squared displacement (MSD) of the photoexcited population profile observed in stroboSCAT as $MSD = \sigma^2(t) - \sigma^2(0)$, where σ is the Gaussian width of the population density profile at time delay t (*13*). Figure 1G compares the MSD for exciton-polarons in $Re_6Se_8Cl_2$ against the MSD for excitons in bilayer WSe_2 and charge carriers in intrinsic monocrystalline Si (*14*), which exhibit some of the best transport among 2D and 3D semiconductors, respectively. We find that the MSD at 1.1 ns in $Re_6Se_8Cl_2$ is 23 times that in bulk WSe_2, 65 times that in bilayer WSe_2, 120 times that in monolayer WSe_2, and almost twice that of electrons in intrinsic Si and the recently reported cubic boron arsenide (*15*, *16*) (table S1). On the basis of the 11-ns polaron lifetime (fig. S5), we estimate that the polaron propagation length in $Re_6Se_8Cl_2$ would exceed 25 μm in the absence of crystal boundaries.

The superlinear behavior of the MSD for $Re_6Se_8Cl_2$ differs from the linear behavior in Si and WSe_2. We fit the MSD to a power law, $MSD \propto t^{\alpha}$ (*11*, *12*, *17*). In the limit of diffusive transport, where scattering lengths are much shorter than the propagation length, $\alpha = 1$. This regime, exemplified by the linear MSD for Si and WSe_2 in Fig. 1F, is observed in virtually all semiconductors beyond the first few femtoseconds after photoexcitation (*18*). In the limit of coherent, ballistic transport (no scattering), $\alpha = 2$, meaning that distance is proportional to time with a slope that defines the velocity. In $Re_6Se_8Cl_2$, we observe quasi-ballistic transport ($\alpha = 1.67 \pm 0.13$) (Fig. 1G) sustained for nanoseconds, until the flake edge is reached. Plotting the same data as distance versus time provides an effective propagation velocity of 2.3 km/s (fig. S6). Monte Carlo simulations reproducing the observed MSD yield an exciton–lattice scattering time of 215 ps, indicating an extraordinary mean free path between scattering events exceeding 1 μm for exciton-polarons in $Re_6Se_8Cl_2$ (Fig. 1H; see figs. S7 and S8 for simulation details and alternative models). These findings reveal that the mechanism for fast and long-range transport in $Re_6Se_8Cl_2$ is efficient shielding from scattering, amply compensating for the large effective mass (and thus low intrinsic velocity) of the polaron. These results are reproducible in multiple flakes of different

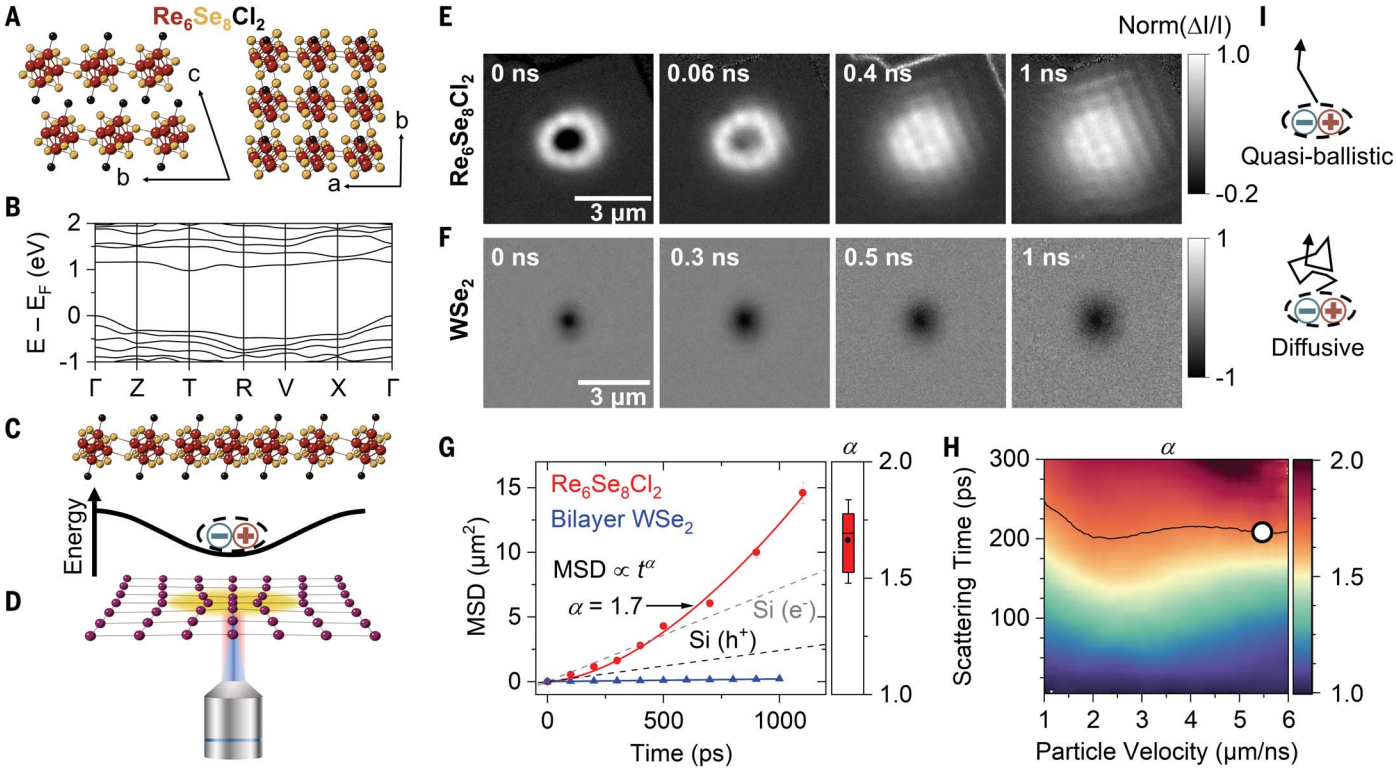


Fig. 1. Imaging exciton transport in $Re_6Se_8Cl_2$. (**A**) Crystal structure of $Re_6Se_8Cl_2$. (**B**) Band structure of $Re_6Se_8Cl_2$ calculated using density functional theory with the Perdew-Burke-Ernzerhof (PBE) exchange correlation functional (*13*). (**C**) Formation of acoustic polarons through a deformation potential interaction. (**D**) Schematic for optical far-field imaging of polaron transport (details in the text and figs. S1 to S3). (**E**) Picosecond stroboSCAT time series displaying exciton (dark contrast) and exciton-polaron (bright contrast) propagation in a 60-nm-thick $Re_6Se_8Cl_2$ flake on glass. Δ*I*/*I* refers to the pump-induced change in backscattered light intensity. (**F**) Exciton propagation in bilayer WSe_2 on glass. (**G**) Mean squared displacement (MSD) of exciton-polarons in $Re_6Se_8Cl_2$ (red), excitons in WSe_2 (blue), and charge carriers in Si (gray and black). Error bars are 1 SD. Only $Re_6Se_8Cl_2$ displays superlinear behavior, indicating superdiffusive transport characterized by the exponent α. The box plot shows the spread of α values across 11 different datasets, indicating mean and median values of 1.67 and 1.7, respectively. (**H**) Monte Carlo simulation of α for different particle velocities (*v*) and scattering times (τ) for our experimental configuration. The circle corresponds to simulation parameters of *v* = 5.5 km/s and τ = 215 ps that reproduce the experimental MSD. The black contour traces α = 1.67. (**I**) Illustration of quasi-ballistic motion of polarons in $Re_6Se_8Cl_2$ compared with diffusive motion for excitons in WSe_2 and other semiconductors. Data acquired at 295 K.

thicknesses (Figs. 1 and 3), for both above-gap and band-edge pump excitation (fig. S9), for pump temporal pulse widths spanning 60 fs to 60 ps (fig. S10), and across a range of fluences explored in detail below (Fig. 3), indicating that neither hot carrier transport (*19*), phonon winds (*20*, *21*), nonlinear recombination (*22*), thermal gradients (*23*), nor strain waves (*24*–*27*) are responsible for the observed behavior (figs. S9 to S12 and table S2).

Excitons in $Re_6Se_8Cl_2$ form exciton-polarons

To confirm that polaron formation is responsible for the observed transport behavior, we tracked the energetic evolution of excitons after photoexcitation and correlated these dynamics to optical transport measurements. Figure 2A displays transient reflectance spectra in the region of the semiconductor band edge, exhibiting a bleach around 1.57 eV and a photoinduced absorption ~90 meV higher in energy. The primary dynamic evolution observed is a spectral redshift on a 1.5-ps timescale that convolves all peaks. Identical dynamics are observed for near-band-edge excitation (fig. S9), ruling out the possibility that electronic thermalization is responsible for the redshift. Tracking the zero-crossing of the transient reflectance profile (dashed line in Fig. 2A) provides a handle on redshift kinetics, plotted in Fig. 2B. We observe an overall 48 meV redshift with an evolution resembling a strongly damped oscillator (red line in Fig. 2B). These dynamics echo those previously observed for (large) polaron formation (*28*–*31*), wherein energetic stabilization occurs over a single vibrational period of the associated lattice deformation. This spectral evolution is responsible for the switch from dark to bright contrast in stroboSCAT in Fig. 1E: The probe at 1.55 eV is initially pre-resonant with the exciton transition (dark contrast) but switches to resonant with the induced polaron absorption after the redshift, generating a bright contrast (additional stroboSCAT datasets at different probe wavelengths are displayed in fig. S13). Below, we discuss measurements of correlated spatio-energetic dynamics, fluence-dependent redshift dynamics, and population saturation that cement our assignment of these redshift dynamics to polaron formation.

Polarons should exhibit correlated spatial and spectral dynamics, given that energetic stabilization associated with polaron formation results in a modification of transport properties. To directly image these correlated dynamics and further support the correspondence between the redshift timescale and the polaron formation timescale, we developed an approach capable of simultaneously resolving spectral and spatial evolution by merging stroboSCAT with transient reflectance microscopy (fig. S2). Figure 2C displays spatially resolved transient spectra at different pump-probe time delays (see also movies S2 and S3). At early times, the transient reflectance signal associated with excitons is concentrated around the pump excitation location at 0 μm. Between 4 and 50 ps, a V-shaped pattern emerges (highlighted with arrows in Fig. 2C), indicating that the spectral redshift associated with polaron formation is correlated to transport away from the excitation location. By 1 ns, only the redshifted polaron spectral signature remains, showing propagation over several micrometers. The correlated spatio-energetic dynamics are a clear indication that

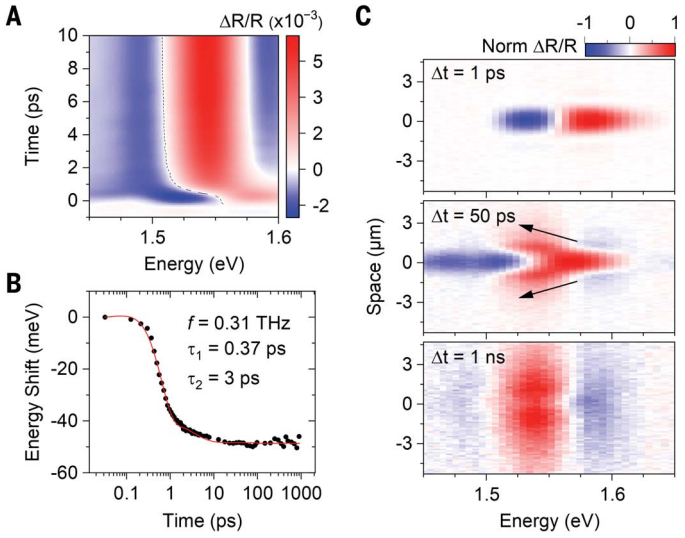


Fig. 2. Spatio-energetic tracking of polaron formation in $Re_6Se_8Cl_2$. (**A**) Transient reflectance spectra near the band edge after 2.41 eV pump excitation at an initial exciton density of 5×10^{17}/cm^3. $\Delta R/R$ refers to the pump-induced change in reflected light intensity. (**B**) Trace of the redshift highlighted with a dashed line in (A). The red line is a fit using a damped oscillator model with a frequency of 0.31 THz and damping time of 0.37 ps, combined with a 3-ps exponential decay. (**C**) Spacetime-transient reflection spectra after 2.41 eV pump excitation at an initial exciton density of 8×10^{18}/cm^3 for three time delays. The correlated redshift and transport dynamics are emphasized with arrows in the middle panel. Movies S2 and S3 display these coupled spatio-energetic dynamics for different excitation fluences. Data acquired at 295 K.

excitons in $Re_6Se_8Cl_2$ become substantially mobile only after they have formed exciton-polarons; bare excitons are effectively immobile, whereas polarons propagate over micrometers. The observed transport dynamics also rule out traditional exciton self-trapping (small polarons), which would reduce the exciton mobility upon polaron formation (*32*). These results illustrate the power of our correlative approach for understanding how polaron formation affects transport.

Finally, we confirmed the formation of polarons in $Re_6Se_8Cl_2$ and experimentally inferred their size by determining the density at which they begin to interact. At high densities, lattice deformations compete to displace the same atoms, resulting in a diminished ability to form polarons (Fig. 3A) (*32*). stroboSCAT data displayed in Fig. 3B shows that as the photoexcitation density increases, the bare exciton signal (dark contrast) begins to dominate over the polaron signal (bright contrast), indicating that polaron formation is suppressed in regions of high excitation density. Figure 3C compares the polaron population (red trace) and bare exciton population (black trace) as a function of excitation fluence (analysis in fig. S14). We observed clear saturation of the polaron population at exciton densities between 0.45×10^{18}/cm^3 and 1.8×10^{18}/cm^3 (highlighted with a blue rectangle). This behavior signals that polarons are overlapping in space, analogous to a Mott transition that prevents further polaron formation (*33*). The onset of polaron saturation is also reflected in a transition from superdiffusive to almost diffusive transport (Fig. 3D), which we attribute to polaron–polaron scattering above the saturation density. The suppression of exciton-to-polaron conversion above saturation is most evident in the spectral dynamics (Fig. 3E and fig. S15), where the redshift time associated with exciton stabilization increases from ~1.5 ps to hundreds of picoseconds. We reproduced these spectral dynamics (right panels of Fig. 3E) with a saturation model accounting for a kinetic blockade and polaron transport away from the excitation area (figs. S15 and S16). According to the Mott criterion (*34*), the polaron interaction radius associated with the observed critical density range of 0.45×10^{18}/cm^3 to 1.8×10^{18}/cm^3 is 2.1 to 3.4 nm, corresponding to three- to five-unit cells in $Re_6Se_8Cl_2$.

Acoustic polarons are responsible for wavelike transport

Large polarons, known to form in materials such as lead halide perovskite semiconductors, have been suggested to partially shield carrier-lattice scattering (*35*). Nevertheless, experiments consistently demonstrate a diffusive transport regime (*10*, *18*, *19*, *36*), with sub-100-fs scattering times that indicate insufficient shielding to switch into the much-desired macroscopic ballistic transport regime. In contrast, the sustained quasi-ballistic behavior observed in $Re_6Se_8Cl_2$ is reminiscent of acoustic polarons, which can form in low-dimensional materials and were theoretically invoked to rationalize the transport properties of polydiacetylene in one dimension (*37*, *38*).

The formation of acoustic polarons is rare, as is our observation of micrometer-scale exciton mean free paths at room temperature. To rationalize this notable behavior in $Re_6Se_8Cl_2$, we use an approximate strong-coupling theory that describes an exciton of mass m coupled to acoustic phonons with a strength quantified by the deformation potential D in two dimensions. The acoustic phonons derive from superatoms of mass M with intercluster vibrational frequency Ω. In two dimensions, the energy to form a circular polaron with radius a and area πa^2 is

$$E = \frac{\hbar^2}{2ma^2} + \frac{1}{2}M\Omega^2\Delta^2\pi a^2 - D\Delta \qquad (1)$$

where Δ is the dimensionless lattice displacement and $\hbar$ is the reduced Planck constant (*39*). In this simple picture, the polaron is only bound if the electron–phonon interaction outweighs the energetic penalties associated with exciton localization and lattice deformation. Minimizing Eq. 1 with respect to the lattice displacement Δ gives the existence criterion

$$\lambda > \lambda_c, \lambda = \frac{mD^2}{2\hbar^2 M\Omega^2} = \frac{D^2}{4JMs^2},$$
$$\lambda_c = {}^{\pi}/_{2} \approx 1.6 \qquad (2)$$

where λ is a dimensionless measure of the exciton–acoustic phonon coupling strength, and in the second equality we have introduced the exciton transfer integral J and the speed of sound $s = \Omega a$. Applying density functional theory to $Re_6Se_8Cl_2$, we calculate D = 4.4 eV (*13*). Taking into account the experimentally inferred electronic band flattening due to intercluster optical phonons, which increases the exciton effective mass from $1.9m_e$ at 0 K to $60m_e$ at 300 K (*7*), we calculate a very small J of 1.5 meV at 300 K (*13*). When combined with the other material parameters (table S3), we find that $\lambda = 7 > \lambda_c$, predicting a strongly bound polaron. Figure 4A plots the coupling strength λ for $Re_6Se_8Cl_2$, monolayer WSe_2, crystalline pentacene, and 2D organic-inorganic halide perovskites. $Re_6Se_8Cl_2$ has a coupling strength λ that is 10 to 1000 times greater than that of the other materials and is thus the only material considered in Fig. 4 that is predicted to exhibit bound acoustic polarons according to the criterion in Eq. 2. Within this theory, the key parameters setting $Re_6Se_8Cl_2$ apart from other 2D and 3D semiconductors is the combination of a quasi-flat electronic band structure at room temperature (small J) and strong exciton–acoustic phonon interactions (large D), yielding a large coupling strength λ and associated strongly bound acoustic polarons. The polaron binding energy is reduced with decreasing temperature owing to the increase in the transfer integral J (*7*, *13*). The estimate in Eq. 2 predicts that the polaron is not bound below ~175 K. Temperature-dependent stroboSCAT experiments (fig. S17) display a drastic reduction in MSD below ~150 K, lending support to our central hypothesis and theory of acoustic polarons. We emphasize that Eq. 2 only provides a qualitative criterion for polaron formation and a detailed understanding of the polaron stability and lifetime at finite temperature requires a more complete theoretical treatment (*13*).

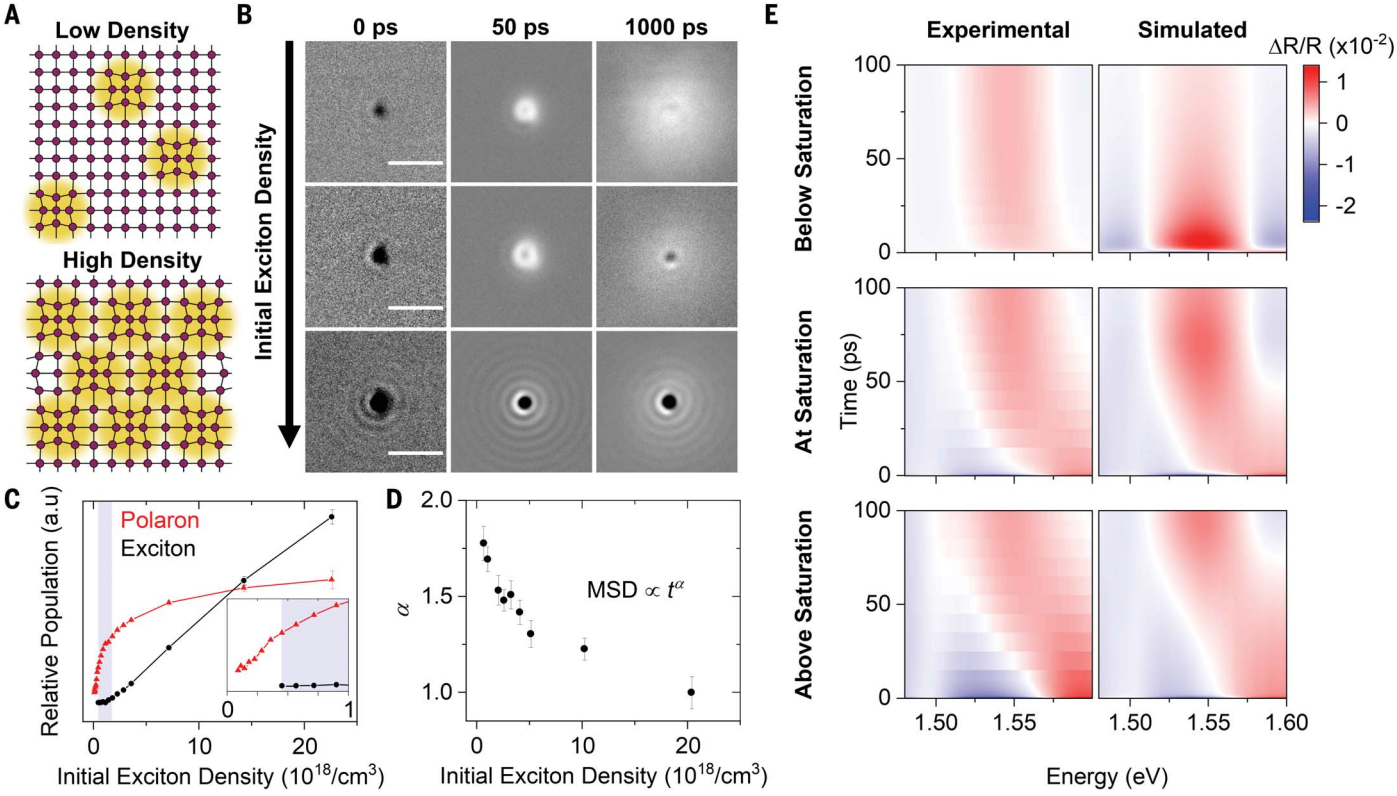


Fig. 3. Determining polaron size in $Re_6Se_8Cl_2$. (**A**) Depiction of polaron blockade at high densities. (**B**) Femtosecond stroboSCAT images at pump/probe energies of 2.41 eV/1.55 eV at initial exciton densities ranging from 1×10^{18}/cm^3 to 20×10^{18}/cm^3 in a 350-nm-thick $Re_6Se_8Cl_2$ flake on glass. Bare excitons are associated with dark contrast, whereas polarons are associated with bright contrast. The rings observed at high exciton densities are diffraction rings. All scale bars are 3 μm. (**C**) Relative populations of polarons and bare excitons at a pump-probe time delay of 5 ps as a function of initial exciton density, indicating saturation of the polaron population. The blue shaded region for exciton densities between 0.45×10^{18}/cm^3 and 1.8×10^{18}/cm^3 indicates the range of critical polaron overlap density. (**D**) MSD exponent, α, as a function of initial exciton density. Error bars are 1 SD. (**E**) Experimental (left) and simulated (right) transient reflectance spectra taken at the center of the focused pump excitation (B) for different initial exciton densities. Simulations are based on a saturation model accounting for exciton-to-polaron conversion and polaron transport away from the excitation spot (*13*). Data acquired at 295 K.

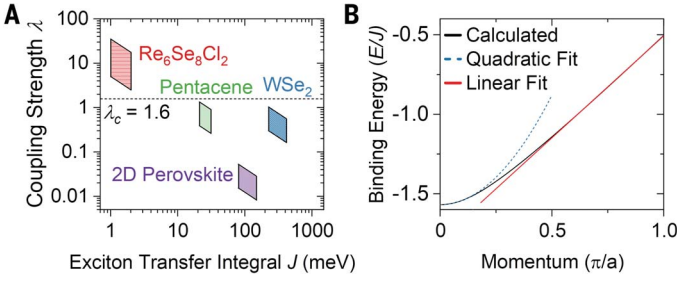


Fig. 4. Energy of an acoustic polaron. (**A**) The coupling strength λ for different systems, presented as a function of their exciton transfer integral *J*. The boxed regions represent the upper and lower bounds for a range of deformation potentials and exciton transfer integrals within 30% of reported data (material properties collected in table S3). The critical value of $\lambda_c = 1.6$ is plotted as a dashed line. (**B**) Calculated polaron dispersion for $Re_6Se_8Cl_2$ (*13*). The band effective mass is estimated by a parabolic fit at the minimum (dashed blue line). The solid red line emphasizes the linearity of the dispersion for higher values of momentum.

Rationalizing the quasi-ballistic dynamics of acoustic polarons requires a more sophisticated quantum mechanical treatment. We generalize the stationary polaron description above and propose a variational wave function for the moving polaron defined by its average crystal momentum (*13*). Energy minimization produces the polaron dispersion shown in Fig. 4B. Near the band bottom, we extract a large effective polaron mass $m^* \approx 200m_e$, a substantial increase from the bare exciton mass of $60m_e$ at 300 K. More importantly, at higher momenta, the polaron inherits the linear dispersion of acoustic phonons—a renormalization evocative of light–matter hybridization to form polaritons (*40*). The linear dispersion of acoustic polarons with a slope below that of other acoustic phonons implies weak scattering because there are no dissipation channels that conserve both energy and momentum (*38*). The polaron is thus predicted to move quasi-ballistically at a speed proportional to the speed of sound of the lattice, consistent with our experimental observations.

Discussion and outlook

We have observed a transport regime mediated by acoustic exciton-polarons in the vdW superatomic semiconductor $Re_6Se_8Cl_2$. Polaron formation shields excitons from scattering with lattice phonons, resulting in quasi-ballistic electronic energy flow over several micrometers within a nanosecond at room temperature. We reveal a very long exciton mean free path of ~1 μm, suggesting the possibility of ballistic excitonic transistors. Our discovery of this regime in a material with weak electronic dispersion provides an alternative to the current paradigm of increasing electronic conjugation to improve transport. Indeed, our model for 2D acoustic polarons suggests that quasi-flat electronic bands and strong electron–phonon interactions can counterintuitively result in

exceptional electronic transport. Beyond 2D superatomic materials such as $Re_6Se_8Cl_2$ and the recently reported graphullerene (*41*), moiré superlattices of 2D semiconductors may provide an interesting testing ground for acoustic polarons. Their superlattice potentials enable tuning electronic bands (*42*, *43*) to achieve values of J down to ~0. Combined with their strong deformation potentials (*44*), these flat bands could yield acoustic polarons with tunable transport properties across a large temperature range. Generalizing wavelike, ultralong-range electronic energy flow in 2D materials could herald an era of essentially lossless nanoelectronics.

REFERENCES AND NOTES

1. A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai, *Nature* **424**, 654–657 (2003).
2. G. D. Scholes *et al*., *Nature* **543**, 647–656 (2017).
3. X. Zhong *et al*., *Nano Lett.* **18**, 1483–1488 (2018).
4. B. Choi *et al*., *J. Am. Chem. Soc.* **140**, 9369–9373 (2018).
5. E. A. Doud *et al*., *Nat. Rev. Mater.* **5**, 371–387 (2020).
6. E. J. Telford *et al*., *Nano Lett.* **20**, 1718–1724 (2020).
7. Q. Li, F. Liu, J. C. Russell, X. Roy, X. Zhu, *J. Chem. Phys.* **152**, 171101 (2020).
8. K. Lee *et al*., *Adv. Mater.* **31**, e1903209 (2019).
9. J. Chen, A. J. Millis, D. R. Reichman, *Phys. Rev. Mater.* **2**, 114801 (2018).
10. M. Delor, H. L. Weaver, Q. Yu, N. S. Ginsberg, *Nat. Mater.* **19**, 56–62 (2020).
11. T. Zhu, J. M. Snaider, L. Yuan, L. Huang, *Annu. Rev. Phys. Chem.* **70**, 219–244 (2019).
12. N. S. Ginsberg, W. A. Tisdale, *Annu. Rev. Phys. Chem.* **71**, 1–30 (2020).
13. Materials and methods are available as supplementary materials.
14. C. Jacoboni, C. Canali, G. Ottaviani, A. Alberigi Quaranta, *Solid-State Electron.* **20**, 77–89 (1977).
15. J. Shin *et al*., *Science* **377**, 437–440 (2022).
16. S. Yue *et al*., *Science* **377**, 433–436 (2022).
17. G. M. Akselrod *et al*., *Nat. Commun.* **5**, 3646 (2014).
18. J. Sung *et al*., *Nat. Phys.* **16**, 171–176 (2020).
19. Z. Guo *et al*., *Science* **356**, 59–62 (2017).
20. M. M. Glazov, *Phys. Rev. B* **100**, 045426 (2019).
21. L. M. Smith *et al*., *Phys. Rev. B* **39**, 1862–1870 (1989).
22. M. Kulig *et al*., *Phys. Rev. Lett.* **120**, 207401 (2018).
23. R. Perea-Causín *et al*., *Nano Lett.* **19**, 7317–7323 (2019).
24. P. Ruello, V. E. Gusev, *Ultrasonics* **56**, 21–35 (2015).
25. J. Rudolph, R. Hey, P. V. Santos, *Phys. Rev. Lett.* **99**, 047602 (2007).
26. E. Baldini *et al*., *Sci. Adv.* **5**, eaax2937 (2019).
27. K. Datta *et al*., *Nat. Photonics* **16**, 242–247 (2022).
28. K. Miyata *et al*., *Sci. Adv.* **3**, e1701217 (2017).
29. F. X. Morrissey, J. G. Mance, A. D. Van Pelt, S. L. Dexheimer, *J. Phys. Condens. Matter* **25**, 144204 (2013).
30. B. Guzelturk *et al*., *Nat. Mater.* **20**, 618–623 (2021).
31. B. Wu *et al*., *Sci. Adv.* **7**, eabd3160 (2021).
32. D. Emin, *Polarons* (Cambridge Univ. Press, 2012).
33. J. M. Frost, L. D. Whalley, A. Walsh, *ACS Energy Lett.* **2**, 2647–2652 (2017).
34. P. P. Edwards, M. J. Sienko, *Phys. Rev. B* **17**, 2575–2581 (1978).
35. H. Zhu *et al*., *Science* **353**, 1409–1413 (2016).
36. Z. Guo, J. S. Manser, Y. Wan, P. V. Kamat, L. Huang, *Nat. Commun.* **6**, 7471–7479 (2015).
37. H.-B. Schüttler, T. Holstein, *Phys. Rev. Lett.* **51**, 2337–2340 (1983).
38. E. G. Wilson, *J. Phys. C Solid State Phys.* **16**, 6739–6755 (1983).
39. G. Whitfield, P. B. Shaw, *Phys. Rev. B* **14**, 3346–3355 (1976).
40. D. L. Mills, E. Burstein, *Rep. Prog. Phys.* **37**, 817–926 (1974).
41. E. Meirzadeh *et al*., *Nature* **613**, 71–76 (2023).
42. L. Wang *et al*., *Nat. Mater.* **19**, 861–866 (2020).
43. F. Wu, T. Lovorn, E. Tutuc, A. H. MacDonald, *Phys. Rev. Lett.* **121**, 026402 (2018).
44. D. Waters *et al*., *ACS Nano* **14**, 7564–7573 (2020).
45. J. Tulyagankhodjaev *et al*., Dataset for "Room temperature wavelike exciton transport in a van der Waals superatomic semiconductor," version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8183811.

ACKNOWLEDGMENTS

We are grateful to P. Batail for bringing this family of materials to our attention and for ensuing productive discussions. We also thank L. E. Brus, X. Zhu, C. Nuckolls, and D. R. Reichman for helpful discussions and D. Xu and J. Baxter for technical help with measurements and analysis. **Funding:** This material is primarily based on work supported by the National Science Foundation (NSF) through the Columbia MRSEC on Precision-Assembled Quantum Materials (PAQM) (DMR-2011738) (M.D., X.R., and T.C.B.) and by the Air Force Office of Scientific Research (AFOSR) under grant number FA9550-22-1-0389 (M.D. and X.R.). stroboSCAT instrument development was supported by the NSF under grant number DMR-2115625 (M.D.). M.D. acknowledges support from the Arnold and Mabel Beckman Foundation through a Beckman Young Investigator award. X.R. and J.Y. acknowledge support from NSF CAREER Award DMR-1751949. J.A.T. was supported by an NSF Graduate Research Fellowship. J.C.R. was supported by a Department of Defense National Defense Science and Engineering Graduate Fellowship. **Author contributions:** J.A.T. and M.D. conceived of and designed the experiments. J.A.T. performed, analyzed, and simulated stroboSCAT and spacetime transient reflection experiments with assistance from A.C.S. J.Y., J.C.R., D.G.C., and X.R. synthesized and characterized $Re_6Se_8Cl_2$ single crystals. J.Y. performed and analyzed heat capacity measurements. J.A.T., M.E.R., J.Y., and H.S. prepared exfoliated samples. P.S. and T.C.B. developed the theory and performed density functional theory calculations and polaron dispersion calculations. M.D. supervised the project. J.A.T., P.S., J.Y., X.R., T.C.B., and M.D. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code necessary to reproduce the figures in the main text and supplementary materials are available on Zenodo (*45*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2698
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S3
References (*46*–*71*)
Movies S1 to S3

Submitted 11 October 2022; accepted 21 September 2023
10.1126/science.adf2698

QUANTUM GASES

Universal scaling of the dynamic BKT transition in quenched 2D Bose gases

Shinichi Sunami[1]*, Vijay Pal Singh[2,3], David Garrick[1], Abel Beregi[1], Adam J. Barker[1], Kathrin Luksch[1], Elliot Bentine[1], Ludwig Mathey[4,5], Christopher J. Foot[1]

The understanding of nonequilibrium dynamics in many-body quantum systems is a fundamental issue in statistical physics. Experiments that probe universal properties of these systems can address such foundational questions. In this study, we report the measurement of universal dynamics triggered by a quench from the superfluid to normal phase across the Berezinskii-Kosterlitz-Thouless transition in a two-dimensional (2D) Bose gas. We reduced the density by splitting the 2D gas in two, realizing a quench across the critical point. The subsequent relaxation dynamics were probed with matter-wave interferometry to measure the local phase fluctuations. We show that the time evolution of both the phase correlation function and vortex density obeys universal scaling laws. This conclusion is supported by classical-field simulations and interpreted by means of real-time renormalization group theory.

The relaxation dynamics of a many-body system that is quenched out of equilibrium display a wide range of scenarios, from simple exponential decay to relaxation through metastable or prethermalized states (*1*, *2*), including phenomena such as pattern formation (*3*), and the absence of thermalization (*4*). Systems that are quenched across a phase transition are particularly intriguing because of their universal self-similar behavior, expected in systems as diverse as superfluid helium (*5*), liquid crystals (*6*), biological cell membranes (*7*), the early universe (*8*), and cold atoms (*9*, *10*). There are numerous theoretical challenges in the treatments of nonequilibrium dynamics [e.g., (*11*)], and this motivates in-depth experimental studies to guide and test theories.

For this purpose, ultracold gases have emerged as a platform of unprecedented control and tunability, serving as quantum simulators for the investigation of many-body dynamics. This approach has led to the observation of Kibble-Zurek (KZ) scaling (*12*–*15*) and universal scaling laws (*16*–*18*) after a quench. Despite theoretical interest (*19*–*21*), universal critical dynamics in nonequilibrium continuous two-dimensional (2D) quantum gases remain elusive because of the lack of precise experimental probes. We show how fluctuations of the many-body 2D system can be probed directly through the extension of local matter-wave interferometry (*22*) similar to that previously used to probe the local phase fluctuations of near-integrable 1D

[1]Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK. [2]Institut für Theoretische Physik, Leibniz Universität Hannover, 30167 Hannover, Germany [3]Quantum Research Centre, Technology Innovation Institute, Abu Dhabi, United Arab Emirates. [4]Zentrum für Optische Quantentechnologien and Institut für Laserphysik, Universität Hamburg, 22761 Hamburg, Germany. [5]The Hamburg Centre for Ultrafast Imaging, Hamburg 22761, Germany.
*Corresponding author. Email: shinichi.sunami@physics.ox.ac.uk

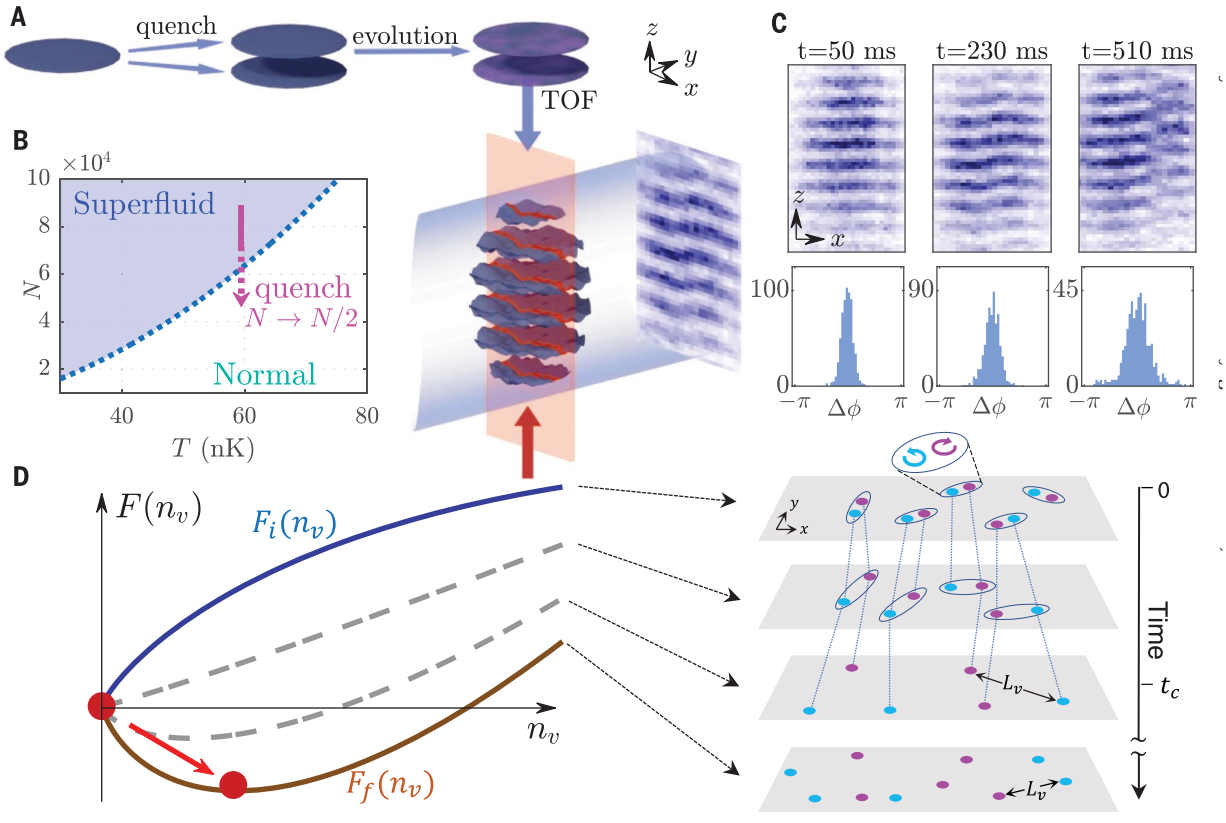


Fig. 1. Observation of nonequilibrium dynamics in 2D Bose gases with matter-wave interferometry. (**A**) A 2D superfluid is split into two daughter clouds, thereby quenching through the BKT transition. The two clouds evolve for time *t* and are released to produce matter-wave interference after a TOF. Local phase fluctuations are observed by optically pumping the slice (red sheet) and then performing absorption imaging. (**B**) Equilibrium phase diagram of trapped 2D Bose gases (*27*). The quench forces the system out of equilibrium toward the normal phase. (**C**) (Top) Examples of interference images. Phase dislocation caused by a vortex is visible in the image at 510 ms. (Bottom) The histograms show the phase differences, $\Delta\phi = \phi(x) - \phi(x')$, at $|x - x'| = 5$ μm, from 45 experimental runs. The decreasing height and increasing width indicate increased phase fluctuations. (**D**) (Left) Free energies $F_i(n_v)$ and $F_f(n_v)$ (continuous lines) in equilibrium for the initial and final conditions of the quench, with their minima indicated by red points (*38*). After the quench, the system undergoes nonequilibrium dynamics (dashed lines) and relaxes toward the state with nonzero free-vortex density, n_v. (Right) Illustration of the dynamics, showing the transition between a scale-invariant phase supported by bound vortex-antivortex pairs and the broken scale-invariant phase characterized by free vortices with the mean vortex-vortex distance L_v, where t_c is the crossover time.

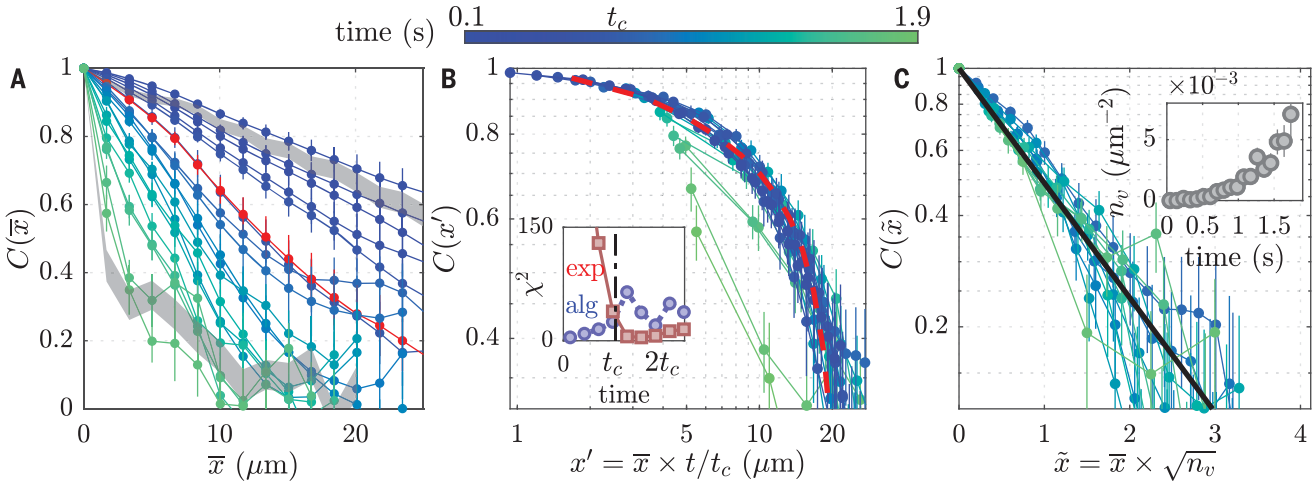


Fig. 2. Nonequilibrium scaling dynamics of the correlation function. (**A**) Relaxation dynamics of the phase correlation function $C(\bar{x})$ measured after the quench for the initial reduced temperature, $\tilde{T} = 0.34$, where $C(\bar{x})$ is an average over 45 realizations or more, and the error bars denote standard error. The data immediately after the critical time, t_c, are indicated by the red curve. For comparison, the measured data for equilibrium 2D gases (*22*) are shown by the gray curves: $\tilde{T} = 0.41$ (top) in the superfluid regime and $\tilde{T} = 0.61$ (bottom) in the normal regime. (**B**) Linear scaling of length $x' = \bar{x}t/t_c$, according to time t/t_c, results in a collapse toward a common curve for time evolution up to $t \sim 1$ s. This universal function is compatible with the expected behavior at the crossover in equilibrium (red dashed line), including the effect of inhomogeneity (*27*). At long times, deviations are observed (green points), indicating the breaking of scale invariance. (Inset) The χ^2 values for the algebraic and exponential fit functions (*27*). (**C**) Rescaling of the distance $\tilde{x} = \bar{x}\sqrt{n_v(t)}$, according to the vortex density $n_v(t)$, results in a collapse of the curves for times $t > 1$ s. Scaling behavior is compatible with an exponential decay (black solid line). (Inset) $n_v(t)$.

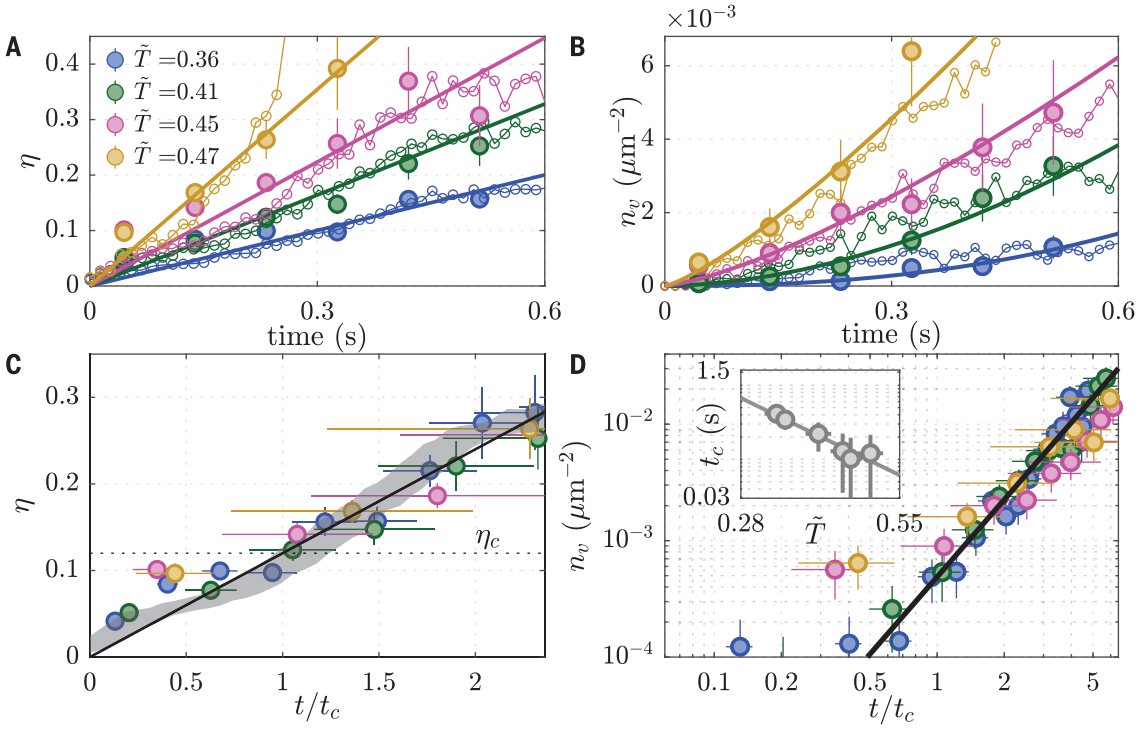


Fig. 3. Universal behavior across the dynamic BKT transition. (**A**) Measured algebraic exponent $\eta(t)$ after the quench at different initial $\tilde{T}$ (filled circles), where the equilibrium critical temperature, $\tilde{T}_c = 0.53$, is crossed by the quench (*22*). Open circles indicate the corresponding simulation results (*27*). Solid lines indicate linear fits to the experimental data, and error bars denote standard fit errors. (**B**) Time evolution of the measured (filled circles) and simulated (open circles) vortex density, $n_v(t)$. The solid lines are power-law fits to the experimental data. Error bars are statistical, given by the square root of observed vortex number. (**C**) Time evolution $\eta(t/t_c)$, scaled according to the $\tilde{T}$-dependent crossover time, t_c. The horizontal error bars arise from the uncertainty in t_c. This universal behavior is used to determine the critical exponent, $\eta_c = 0.13(1)$ (horizontal dotted line), at $t/t_c = 1$. The gray shaded curve indicates the simulation result at $\tilde{T} \sim 0.4$, and the solid line is a guide for the eye. (**D**) Scaled time evolution $n_v(t/t_c)$, plotted on a log-log scale, displays a universal growth after t_c. Black solid line is the fit with power law $n_v \propto t^{2\nu}$, which yields $\nu = 1.1(1)$. (Inset) Dependence of t_c on $\tilde{T}$, with a solid line as a guide for the eye.

systems (*1*, *2*). A feature of 2D systems is that, unlike 1D systems, they exhibit phase transitions with associated critical phenomena. An especially interesting case is that of the critical dynamics across the Berezinskii-Kosterlitz-Thouless (BKT) transition (*23*–*25*), when the system is quenched from the superfluid to the normal phase, the opposite of commonly used quenches that move from the disordered to ordered phase. Real-time renormalization-group (RG) theory and truncated Wigner simulations (*19*) predict that the relaxation occurs

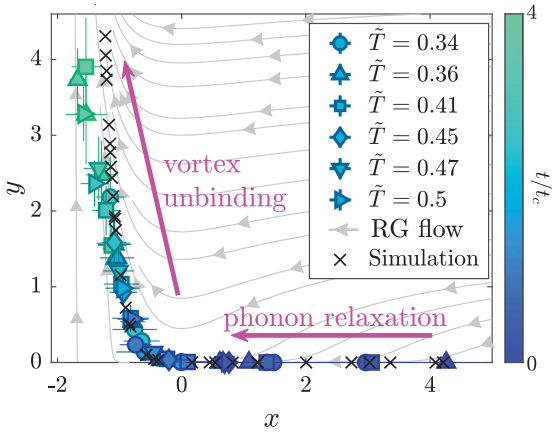


Fig. 4. Real-time renormalization group flow and measurements. Gray lines with arrows are the RG flow of the parameters $x = 1/(2\eta) - 1/(2\eta_c)$ and $y = \sqrt{2}\pi g$ [Eqs. 1 and 2 and (*27*)]. These results are compared with the experimental data for six different initial temperatures, with the time scaled by the corresponding $\tilde{T}$-dependent crossover time, t_c. Error bars denote standard errors propagated from the values for η and n_v. The results from a numerical simulation at $\tilde{T} = 0.47$ are shown as black crosses, where the deviation from experimental data at small x is attributed to the slow trap-induced heating (*27*).

through a reverse-Kibble-Zurek–type mechanism in which delayed vortex proliferation results in a metastable supercritical phase. For equilibrium systems, the BKT transition is driven by the unbinding of vortex-antivortex pairs (*22*, *26*), underscoring the topological nature of the transition. This transition is characterized by a change of the functional form of the correlation function from a power law deep in the superfluid regime, $g_1(\mathbf{r}, \mathbf{r'}) = \langle \Psi(\mathbf{r})^\dagger \Psi(\mathbf{r'}) \rangle \propto |\mathbf{r} - \mathbf{r'}|^{-\eta}$, to exponential deep in the normal regime, $g_1(\mathbf{r}, \mathbf{r'}) \propto e^{-|\mathbf{r}-\mathbf{r'}|/r_0}$, where $\Psi(\mathbf{r})$ is the bosonic field operator at position vector $\mathbf{r}$, η is the algebraic exponent, and r_0 is a correlation length. A similar change of the functional form of the single-particle correlation function is also expected for the dynamical BKT transition. However, this change is expected to occur as a smooth crossover between the two phases (*19*, *27*).

Here, we studied the critical dynamics across the BKT transition by quenching a 2D Bose gas from the superfluid to the normal phase by splitting it in two. Using spatially resolved matter-wave interferometry, we measured the first-order correlation function and vortex density to analyze their relaxation dynamics. We found that relaxation occurs through a two-step process involving phonon relaxation followed by dynamical vortex proliferation. We demonstrated universal scaling laws for the algebraic exponent and vortex density by performing measurements at different initial conditions. Both real-time RG theory (*19*, *28*) and classical-field simulations are in good agreement with our measurements.

Matter-wave interferometry after a quench

Our experiments began with a single, pancake-shaped quasi-2D Bose gas in the superfluid regime, consisting of $N \approx 9 \times 10^4$ atoms of ^{87}Rb at reduced temperatures in the range $\tilde{T} = 0.34 - 0.5$; here, $\tilde{T} \equiv T/T_0$ is the ratio of the initial temperature, T, and the critical temperature $T_0 = \sqrt{6N}(\hbar\omega_r/\pi k_B) \approx 120$ nK of a non-interacting trapped gas (*29*), where $\omega_r/2\pi = 11$ Hz is the radial trapping frequency, $\hbar$ is the reduced Planck constant, and k_B is the Boltzmann constant. The quench is implemented by a rapid splitting of the system in a multiple-RF–dressed potential (*27*, *30*–*32*) (Fig. 1A), which results in a pair of decoupled clouds, each with atom number $N' = N/2$, trapped in the two minima of a double-well potential. Each well has a vertical-trap frequency of $\omega_z/2\pi = 1$ kHz, hence the dimensionless 2D interaction strength is $\tilde{g} = \sqrt{8\pi}a_s/\ell_0 = 0.076$ (*27*), where a_s is the 3D s-wave scattering length and $\ell_0 = \sqrt{\hbar/(m\omega_z)}$ is the harmonic oscillator length along z for an atom of mass m. The initial reduced temperature, $\tilde{T}$, is chosen to be close to the equilibrium critical point at $\tilde{T}_{c,eq} = 0.53$ so that the value after splitting, $\tilde{T}'/\tilde{T} = 1.67$, corresponds to the normal phase (*27*). To investigate the dynamics, we let each cloud evolve independently for time t before performing a time-of-flight (TOF) expansion of $t_{TOF} = 16$ ms, after which we detected the matter-wave interference that encodes the in situ relative phase fluctuation, $\phi(x)$, of two clouds along a line that goes through the center of the cloud. Interference images and histograms of spatial phase fluctuations $\Delta\phi$ show stronger fluctuations at long evolution times (Fig. 1C). The dynamics across the BKT transition are expected to be scale invariant until the bound vortex-antivortex pairs dissociate to disrupt the phase coherence (Fig. 1D).

Relaxation dynamics

To analyze the relaxation dynamics, we used the interference pattern to determine both the correlation function and the vortex density (*22*). The phase correlation of the system between two points at locations $x - \bar{x}/2$ and $x + \bar{x}/2$, with spatial separation $\bar{x}$, is determined by $C(x - \bar{x}/2, x + \bar{x}/2) = \mathrm{Re}\left[\frac{1}{N_r}\sum_{j=1}^{N_r} e^{i\left[\phi_j(x-\bar{x}/2) - \phi_j(x+\bar{x}/2)\right]}\right]$, where the index, j, runs over $N_r = 45$ experimental repeats; $C(x - \bar{x}/2, x + \bar{x}/2) = 1$ indicates perfect correlation of phases, whereas $C(x - \bar{x}/2, x + \bar{x}/2) = 0$ denotes uncorrelated phases. For quantitative analysis, we average $C(x - \bar{x}/2, x + \bar{x}/2)$ over x within the region of clear interference fringes to obtain $C(\bar{x})$ (*33*). In (*22*), $C(\bar{x})$ of 2D Bose gases at equilibrium was found to decay exponentially in the normal regime, whereas in the superfluid regime it was found to decay algebraically with spatially varying exponent for inhomogeneous 2D gases in the superfluid regime, which is in agreement with the prediction based on local correlation approximation in (*34*).

In Fig. 2A, we show the time evolution of $C(\bar{x})$ after the quench at $t = 0$. Initially, there is almost no spatial correlation decay because the two clouds have nearly identical phases; their phases decouple within ~100 ms (*27*). After this, we observe a temporal decay of $C(\bar{x})$ for all $\bar{x}$. At longer times, we observe rapid spatial decay of correlation, indicating vanishing coherence at large distances. To determine the nature of this dynamic transition, we fit $C(\bar{x})$ with the algebraic and exponential functions that are used to characterize the equilibrium BKT transition (*22*). As the system relaxes from the initial superfluid state to the normal state, it is expected that $C(\bar{x})$ will evolve smoothly from algebraic to exponential scaling in the short- and long-time limits, respectively. At intermediate times, both power-law and exponential fitting can be used to extract physical properties of the nonequilibrium state; within the real-time RG picture, nonequilibrium systems in the crossover region are away from the fixed point and there is no direct correspondence to equilibrium systems (*19*, *35*). Indeed, at short and intermediate times, the spatial decay of the correlation function is compatible with algebraic scaling, including the effect of inhomogeneity of the system, and with exponential scaling for long times (*27*). We identified the crossover time, t_c, as the time at which the correlation function becomes better described by exponential scaling rather than algebraic (Fig. 2B, inset). Even after t_c, the χ^2 values for the power-law model remain small, supporting the description of the system with the power-law exponent η beyond t_c (*27*).

The dynamic BKT transition is expected to have self-similar dynamics with a length scale that depends linearly on time (*19*, *36*, *37*). Motivated by this, we have plotted the correlation functions with rescaled length $x' = \bar{x}t/t_c$, using $t_c \sim 0.5$ s (Fig. 2B). This shows convincingly that, except for long times, the fluctuations in the system depend only on the rescaled parameter x' through a universal function, which we found to be close to the expected function at the equilibrium BKT crossover (Fig. 2B). We found the same behavior independent of the initial condition (temperature) of the system, demonstrating the robustness of the scale-invariant behavior near the critical point (fig. S4).

At long times, scale invariance is broken by vortex excitations, which results in an emergent

length scale, $r_0 \approx 1/\sqrt{n_v}$, where n_v is the vortex density (*27*). To demonstrate this, we plotted the correlation function at long times against the rescaled distance $\tilde{x} = \bar{x}\sqrt{n_v}$ (Fig. 2C) (*37*). We obtained n_v from the occurrence of sudden jumps of phases, which indicate the presence of a vortex core (*22*, *26*, *27*). These transformed correlation functions are nearly time independent, showing that the system can be characterized by the vortex density deep in the normal regime.

Varying the initial conditions

Having verified the behavior of the dynamic BKT transition, we next analyzed its universal characteristics by varying $\tilde{T}$. The time evolution of the algebraic exponent, η, determined by using an algebraic fit to $C(\bar{x})$, exhibits a linear increase in which the increase is faster for higher $\tilde{T}$ (Fig. 3A). This shows that the dynamics are accelerated at higher $\tilde{T}$ and that the system quickly crosses over to the normal phase. This is also reflected in the measurements of the vortex density, n_v, showing a faster growth at higher $\tilde{T}$ (Fig. 3B). We found that the vortex growth follows a power-law scaling as expected from the RG predictions (fig. S8). We compared the measurements of η and n_v with the corresponding results of classical-field simulations, which give consistent dynamics (Fig. 3, A and B).

To confirm universal scaling, η and n_v have been plotted as a function of scaled time, t/t_c (Fig. 3, C and D). The time evolutions for various initial values of $\tilde{T}$ collapse onto a single curve, showing the robustness of dynamical scaling. We found a linear increase of η across $t = t_c$. In equilibrium theory, η scales approximately linearly with temperature in the superfluid regime (i.e., $\eta \propto T/4T_{\mathrm{BKT}}$) (*25*), thus connecting the temperature scale with phase fluctuations. From this linear estimate, we obtained the critical exponent $\eta_c = 0.13(1)$ at $t/t_c = 1$, close to the value $\eta_c = 0.17(3)$ found for a finite-size equilibrium system with similar experimental parameters (*22*) and different from the value of $\eta_{\mathrm{BKT}} = 0.25$ for the equilibrium BKT transition in the thermodynamic limit. The linear increase of η after $t/t_c = 1$ is a precursor of nonequilibrium superheated superfluid (*28*) as a consequence of a delayed vortex proliferation. The vortex growth after $t/t_c = 1$ exhibited universal power-law scaling $n_v \propto t^{2\nu}$, with $\nu \sim 1$; this agrees with the RG prediction, as shown in the next section.

Comparison to renormalization-group theory

We next compared the experimental results with predictions based on the real-time RG equations (*19*, *28*). These equations describe the time evolution of parameters characterizing the system from arbitrary nonequilibrium states flowing toward fixed points, which represent possible equilibrium states. For the dynamic BKT transition, the real-time RG equations are (*19*, *27*, *28*)

$$\frac{dg}{dt} = \left(2 - \frac{1}{2\eta}\right)\frac{g}{t} \tag{1}$$

$$\frac{d\eta}{dt} = \frac{\pi g^2}{16\eta t} + \gamma \tag{2}$$

where the vortex fugacity, g, is related to the vortex density by $n_v(t) = n\tilde{g}\exp[(2\ln(g/2))/(2 - 1/(2n))]$. n is the mean density within the region of interest used for fringe analysis, and $\tilde{g}$ is the interaction strength as defined previously; these parameters also determine the healing length, $\xi = 1/\sqrt{n\tilde{g}}$, characterizing the length scale of the vortex cores (*27*, *38*). This RG flow derives from the dynamical sine-Gordon model, serving as a dual model for describing the BKT transition; we have added a phenomenological heating term, γ, to account for the slow, trap-induced heating of the system (*27*). For $\eta < 1/4$, the fugacity is strongly suppressed, resulting in a linear dispersion, $\omega_k = ck$. As η increases in time, the vortex fugacity becomes relevant and increases, resulting in a dispersion, $\omega_k = c\sqrt{k^2 + 1/r_0^2(t)}$. As we have argued and demonstrated in the previous sections, this is indeed supported by the two-step scaling behavior. Furthermore, at long times, we have $1/(2\eta) \ll 2$, therefore $n_v \propto g \propto t^2$ (Fig. 3D).

In Fig. 4, the experimental observations are plotted together with the RG flow diagram of Eqs. 1 and 2. For this representation, we defined $x = 1/(2\eta) - 1/(2\eta_c)$ and $y = \sqrt{2}\pi g$. This ensures that $x_c = 0$ at $\eta = \eta_c$ independent of system sizes, where $\eta_c = 0.13(1)$ for our finite-sized system, and $\eta_c = \eta_{\mathrm{BKT}} = 1/4$ is the theoretical prediction in the thermodynamic limit. Our results follow a universal trajectory in the flow diagram. The quenched system begins at large x, where vortex excitations are suppressed and fugacity is small. Later on, nonequilibrium phonon creation drives the system toward smaller x, however still with suppressed y. As the system approaches the critical point $x_c = 0$, the onset of vortex excitation drives the transition.

Discussion and outlook

Our work provides a comprehensive understanding of nonequilibrium dynamics across the BKT transition. The experimental measurements support the real-time RG picture of universality out of equilibrium, indicating that it is an excellent starting point for the theoretical study of a wide range of many-body dynamics within the framework of RG. The results also show that our matter-wave interference technique is ideally suited for further in-depth investigation of universal dynamics in 2D systems, such as the Kibble-Zurek scaling (*8*) and nonthermal fixed points (*39*).

REFERENCES AND NOTES

1. T. Langen *et al.*, *Science* **348**, 207–211 (2015).
2. T. Schweigler *et al.*, *Nature* **545**, 323–326 (2017).
3. H. P. Zahn *et al.*, *Phys. Rev. X* **12**, 021014 (2022).
4. T. Kinoshita, T. Wenger, D. S. Weiss, *Nature* **440**, 900–903 (2006).
5. W. H. Zurek, *Nature* **317**, 505–508 (1985).
6. I. Chuang, R. Durrer, N. Turok, B. Yurke, *Science* **251**, 1336–1342 (1991).
7. S. L. Veatch, O. Soubias, S. L. Keller, K. Gawrisch, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17650–17655 (2007).
8. T. W. B. Kibble, *J. Phys. A Math. Theor.* **9**, 1387 (1976).
9. Q. Zhou, T.-L. Ho, *Phys. Rev. Lett.* **105**, 245702 (2010).
10. A. Polkovnikov, K. Sengupta, A. Silva, M. Vengalattore, *Rev. Mod. Phys.* **83**, 863–883 (2011).
11. J. Eisert, M. Friesdorf, C. Gogolin, *Nat. Phys.* **11**, 124–130 (2015).
12. N. Navon, A. L. Gaunt, R. P. Smith, Z. Hadzibabic, *Science* **347**, 167–170 (2015).
13. L. W. Clark, L. Feng, C. Chin, *Science* **354**, 606–610 (2016).
14. J. Beugnon, N. Navon, *J. Phys. At. Mol. Opt. Phys.* **50**, 022002 (2017).
15. A. Keesling *et al.*, *Nature* **568**, 207–211 (2019).
16. M. Prüfer *et al.*, *Nature* **563**, 217–220 (2018).
17. S. Erne, R. Bücker, T. Gasenzer, J. Berges, J. Schmiedmayer, *Nature* **563**, 225–229 (2018).
18. J. A. P. Glidden *et al.*, *Nat. Phys.* **17**, 457–461 (2021).
19. L. Mathey, A. Polkovnikov, *Phys. Rev. A* **81**, 033605 (2010).
20. H.-C. Chu, G. A. Williams, *Phys. Rev. Lett.* **86**, 2585–2588 (2001).
21. A. J. Groszek, P. Comaron, N. P. Proukakis, T. P. Billam, *Phys. Rev. Res.* **3**, 013212 (2021).
22. S. Sunami *et al.*, *Phys. Rev. Lett.* **128**, 250402 (2022).
23. V. Berezinskiĭ, *Sov. J. Exp. Theor. Phys.* **34**, 610 (1972).
24. J. M. Kosterlitz, D. J. Thouless, *J. Phys. C Solid State Phys.* **6**, 1181–1203 (1973).
25. D. R. Nelson, J. M. Kosterlitz, *Phys. Rev. Lett.* **39**, 1201–1205 (1977).
26. Z. Hadzibabic, P. Krüger, M. Cheneau, B. Battelier, J. Dalibard, *Nature* **441**, 1118–1121 (2006).
27. Materials and methods are available as supplementary materials.
28. L. Mathey, K. J. Günter, J. Dalibard, A. Polkovnikov, *Phys. Rev. A* **95**, 053630 (2017).
29. For the range of parameters used in this work, the ratio T/T_0 has a one-to-one mapping to the peak phase-space density of the trapped gas and sufficiently represents the effect of the quench; see (*27*).
30. A. J. Barker *et al.*, *New J. Phys.* **22**, 103040 (2020).
31. T. L. Harte *et al.*, *Phys. Rev. A* **97**, 013616 (2018).
32. A. J. Barker *et al.*, *J. Phys. At. Mol. Opt. Phys.* **53**, 155001 (2020).
33. We analyzed the fringes within a region of interest corresponding to 80% of the Thomas-Fermi radius of the quasicondensate. As shown in fig. S3, this is where the interference fringes are clear and allow reliable measurements.
34. I. Boettcher, M. Holzmann, *Phys. Rev. A* **94**, 011602 (2016).
35. In fact, a combined fitting function can be used that has power-law behavior at short distances and exponential scaling at long distances, where the exponential length scale can be infinite.
36. P. Comaron, F. Larcher, F. Dalfovo, N. P. Proukakis, *Phys. Rev. A* **100**, 033618 (2019).
37. To motivate the scaling regimes, we considered a linear spectrum, $\omega_k = ck$. With this dispersion, dynamical phase factors of the form $\exp(i\omega_k t)$ are kept invariant if a $z = 1$ scaling is applied that leaves ckt invariant. Furthermore, for a spectrum that includes an additional length, r_0, such as $\omega_k = c\sqrt{k^2 + 1/r_0^2(t)}$, for long times, only the modes with $\omega \approx c/r_0(t)$ contribute to the dynamics. Here the scaling is replaced by keeping $ct/r_0(t)$ fixed, as it was demonstrated for the long-time dynamics in Fig. 2C. As discussed in the main text and shown in Fig. 4, the length scale $r_0(t)$ is related to the vortex fugacity increasing at long times.
38. X. G. Wen, *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons* (Oxford Univ. Press, 2010).
39. J. Schole, B. Nowak, T. Gasenzer, *Phys. Rev. A* **86**, 013624 (2012).
40. S. Sunami *et al.*, Dataset for "Universal scaling of the dynamic BKT transition in quenched 2D Bose gases," version 3, Zenodo (2023); https://zenodo.org/records/8385524.

ACKNOWLEDGMENTS

We acknowledge discussions with J. Okamoto on theoretical analysis and thank J. Chalker for comments on our manuscript.

Funding: The experimental work was supported by the EPSRC grant reference EP/S013105/1. S.S. acknowledges the Murata Overseas Scholarship Foundation, the Ezoe Memorial Recruit Foundation, the Daishin Foundation, and St. Hilda's College, Oxford, for financial support. D.G., A.B., A.J.B., and K.L. thank the EPSRC for doctoral studentships. L.M. acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG) in the framework of SFB 925 – project ID 170620586 and the excellence cluster "Advanced Imaging of Matter" – EXC 2056 – project ID 390715994. V.P.S. acknowledges funding by the Cluster of Excellence "QuantumFrontiers" – EXC 2123 – project ID 390837967. **Author contributions:** S.S. performed the experiments and data analysis. V.P.S. and L.M. developed numerical and analytical models and contributed to the interpretation of our experimental data. S.S. and V.P.S. wrote the manuscript. L.M. and C.J.F. supervised the project. All authors contributed to the discussion and interpretation of our results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data presented in this paper and numerical simulation codes are available at Zenodo (*40*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq6753
Materials and Methods
Figs. S1 to S9
References (*41*–*54*)

Submitted 11 May 2022; accepted 16 September 2023
10.1126/science.abq6753

SOLID STATE PHYSICS

Giant lattice softening at a Lifshitz transition in Sr_2RuO_4

H. M. L. Noad[1]*, K. Ishida[1]†, Y.-S. Li[1], E. Gati[1], V. Stangier[2], N. Kikugawa[3], D. A. Sokolov[1], M. Nicklas[1], B. Kim[4,5], I. I. Mazin[6,7], M. Garst[8,9], J. Schmalian[2,9]*, A. P. Mackenzie[1,10]*, C. W. Hicks[1,11]*

The interplay of electronic and structural degrees of freedom in solids is a topic of intense research. More than 60 years ago, Lifshitz discussed a counterintuitive possibility: lattice softening driven by conduction electrons at topological Fermi surface transitions. The effect that he predicted, however, was small and has not been convincingly observed. Using a piezo-based uniaxial pressure cell to tune the ultraclean metal strontium ruthenate while measuring the stress-strain relationship, we reveal a huge softening of the Young's modulus at a Lifshitz transition of a two-dimensional Fermi surface and show that it is indeed driven entirely by the conduction electrons of the relevant energy band.

The coupling between elastic and electronic degrees of freedom is crucial to determining the phase diagrams of correlated electron systems, such as those displaying electronic nematicity, in which conduction electrons develop anisotropies (*1*). However, there is always a "chicken and egg" question: Does the lattice drive or respond to the conduction electron physics (*2*, *3*)? Here, we approached the entanglement of electronic and structural degrees of freedom using a different method from those most commonly used. We studied the stress-strain relationship of the quasi–two-dimensional (2D) correlated metal strontium ruthenate (Sr_2RuO_4) as it was tuned through a saddle point Lifshitz transition (*4*–*6*) in which the Fermi surface topology changes and the Fermi level crosses a Van Hove singularity (VHS) (*7*). By combining direct stress-strain measurements with experimentally determined entropy data across the same transition, we have demonstrated the existence of an unexpectedly large softening of the lattice driven entirely by conduction electrons. The possibility of such effects was discussed theoretically by Lifshitz himself >60 years ago, but their size was estimated to be extremely small (*4*). By contrast, here, we measured a large effect that is in principle singular, i.e., capable of introducing a lattice instability in the $T \rightarrow 0$ limit if not cut off by a phase transition to some other form of order. We discuss our results in the framework of quantum critical elasticity and show that superconductivity is a natural way of cutting off quantum critical lattice softening.

Young's modulus dips across a stress-tuned Lifshitz transition in Sr_2RuO_4

The material platform for our experiments, Sr_2RuO_4, has attracted considerable attention both as an unconventional superconductor (*8*–*12*) and as a benchmark 2D Fermi liquid (*13*, *14*). It is one of the cleanest correlated electron materials known. The best single crystals have residual resistivities of ≈50 nΩ cm, corresponding to electron mean free paths of 2 μm or more (*15*). The Fermi surface of Sr_2RuO_4 consists of three cylinders, commonly referred to as α, β, and γ (*14*). Electronic correlations lead to a substantial mass renormalization over the values predicted in independent-electron band structure calculations, meaning that ur… ambient conditions, the Fermi energy (E_F)… the γ band sits only 14 meV below a saddle point VHS at the M point of the tetragonal Brillouin zone (*16*). It is possible to tune this VHS through E_F by applying a pressure of 0.7 ± 0.1 GPa along the [100] crystalline direction (*17*, *18*), profoundly affecting the electronic properties. For example, the superconducting transition temperature is enhanced by a factor of 2.4 from its ambient pressure value to 3.5 K (*19*), and the temperature dependence of the resistivity undergoes a large change from the conventional Fermi liquid T^2 dependence seen at higher and lower stress (*5*).

To investigate the consequences of this Lifshitz transition on the lattice stiffness, we used a custom apparatus in which both uniaxial stress σ and strain ϵ can be monitored simultaneously (*20*), allowing measurement of the Young's modulus (fig. S1). To maximize the quantitative accuracy of the data, samples were milled into a necked shape, as shown in the inset of Fig. 1A, using a Xe plasma source focused ion beam. The end tabs were then

[1]Max Planck Institute for Chemical Physics of Solids, 01187 Dresden, Germany. [2]Institut für Theorie der Kondensierten Materie, Karlsruher Institut für Technologie, 76131 Karlsruhe, Germany. [3]National Institute for Materials Science, Tsukuba, Ibaraki 305-0003, Japan. [4]Department of Physics, Kunsan National University, Gunsan 54150, Korea. [5]Department of Physics, Kyungpook National University, Daegu 41566, Korea. [6]Department of Physics and Astronomy, George Mason University, Fairfax, VA 22030, USA. [7]Quantum Science and Engineering Center, George Mason University, Fairfax, VA 22030, USA. [8]Institut für Theoretische Festkörperphysik, Karlsruher Institut für Technologie, 76131 Karlsruhe, Germany. [9]Institut für QuantenMaterialien und Technologien, Karlsruher Institut für Technologie, 76131 Karlsruhe, Germany. [10]Scottish Universities Physics Alliance, School of Physics and Astronomy, University of St Andrews, St Andrews KY16 9SS, UK. [11]School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK.
***Corresponding author. Email: hilary.noad@cpfs.mpg.de (H.M.L.N.); joerg.schmalian@kit.edu (J.S.); andy.mackenzie@cpfs.mpg.de (A.P.M.); c.hicks.1@bham.ac.uk (C.W.H.)**
†Present address: Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan.

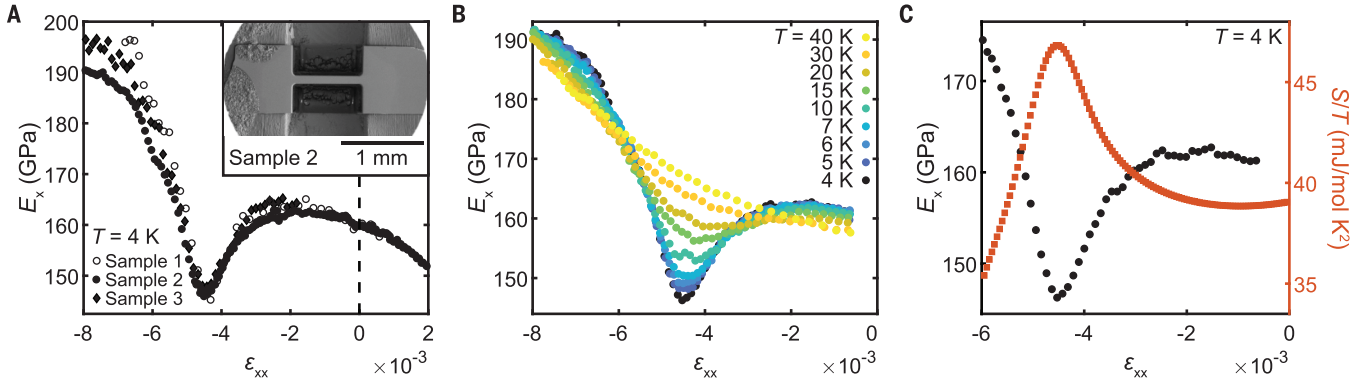


Fig. 1. The Young's modulus E_x of Sr_2RuO_4 measured across a stress-tuned Lifshitz transition. (**A**) E_x as a function of strain ε_{xx} measured at 4 K on three samples. Inset: Scanning electron micrograph of sample 2. (**B**) E_x versus ε_{xx} at a series of temperatures measured on sample 2. (**C**) E_x at 4 K taken from the temperature series in (B) (black) together with the entropy S/T extracted from elastocaloric data from a separate sample at 4 K (orange) plotted as a function of ε_{xx}. The elastocaloric data are from (*23*).

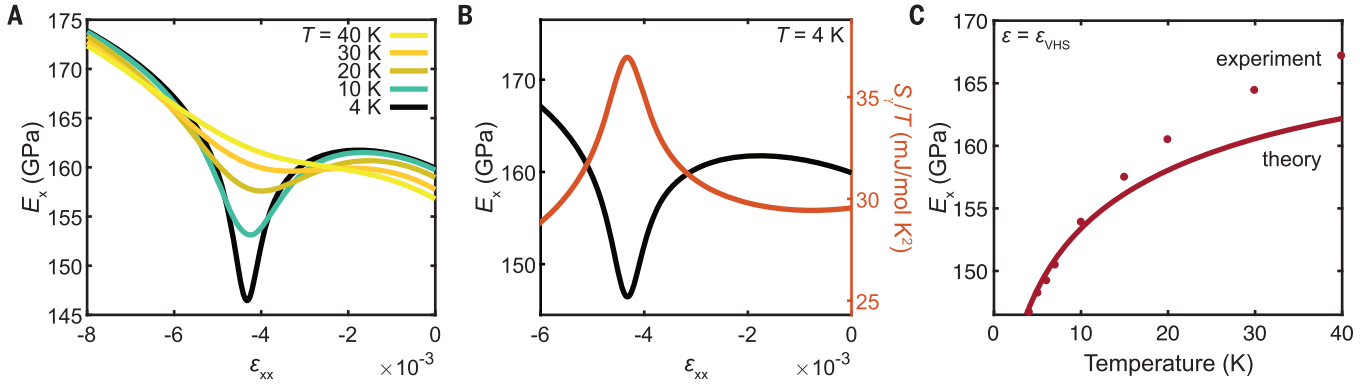


Fig. 2. A simple model quantitatively reproduces key experimental features. (**A**) Calculated E_x versus ε_{xx} at selected temperatures. (**B**) E_x and S_γ/T versus ε_{xx} at 4 K, where S_γ is the calculated contribution to the entropy of the γ band. (**C**) E_x versus temperature calculated at the Van Hove strain (solid line) and corresponding experimental data (filled circles) extracted from the temperature series shown in Fig. 1B. The averaging window for the data is a strain range of 3×10^{-4}. For details on the model, see (*22*).

embedded in epoxy, which acts as a conformal layer through which large forces can be transmitted to the brittle samples. The necking creates a rapid crossover between low- and high-stress regions of the sample, which is important for resolving fine features in the stress-strain relationship (*21*). Full details on how we extracted the Young's modulus and strain of the sample, including an examination of possible systematic errors that could affect the analysis, can be found in (*22*) and in figs. S2 to S5.

Our core result is shown in Fig. 1A: the differential Young's modulus $E_x = d\sigma_{xx}/d\varepsilon_{xx}$ as a function of strain at 4 K. To demonstrate repeatability, data from three samples are shown. Samples 2 and 3 had a higher aspect ratio (*22*) and are therefore expected to yield more accurate data. Each sample was in a different stress cell. The force calibrations of the cells were refined using the known Lifshitz stress of −0.7 GPa (*18*), where negative values denote compression. At the Lifshitz transition strain of $\varepsilon_{VHS} = -(0.45 \pm 0.05)\%$, E_x is seen to drop to ~146 GPa, and then beyond the transition to increase to ~200 GPa. In other words, contrary to our naïve expectation, the dip in E_x is not a small effect; the softening of the lattice at the Lifshitz transition is between 10 and 15% depending on the definition used for the background value. Under tensile strain, E_x again decreases quite rapidly. A Lifshitz transition under tensile strain equivalent to that under compressive strain would be expected, and this decrease in E_x is most likely caused by the approach to that transition. A second notable aspect of the lattice softening at the Lifshitz strain is its strong temperature dependence, shown in Fig. 1B. At 40 K, the dip is barely resolved. As the temperature is lowered, it sharpens and deepens, with a substantial change observable even between 5 and 4 K.

Elastic constants are second derivatives of the free energy, $C = \frac{\partial^2 F}{\partial \varepsilon^2}$. $F = U - TS$, where U is the internal energy and S is entropy. The intuitive expectation might be that the valence band contributions to U completely dominate F and thus also the elastic moduli. However, the generally valid expression for the free energy $F(T,\varepsilon) = E_0(\varepsilon) - \int_0^T S(T',\varepsilon)dT'$, where E_0 is the ground state energy, reveals that any temperature dependence of elastic constants derives from changes in the entropy even if $E_0(\varepsilon)$ may dominate the elastic moduli. Our Young's modulus data have a strong temperature dependence (Fig. 1B), so a link between the Young's modulus and a strain-dependent entropy would be expected. In Fig. 1C, we compare the entropy obtained from a recent study of the elastocaloric effect on Sr_2RuO_4 (*23*) at 4 K with the Young's modulus data at the same temperature. The strong correlation between the two leads to the conclusion that the key physics that we observe are driven by a conduction band. At this low temperature, the phonon contribution to the entropy in Sr_2RuO_4 is negligible (*24*) and the valence band contribution even more so. All that would be observed is conduction band entropy because of the density of states at the Fermi level.

Modeling lattice softening

To understand the observed behavior quantitatively, we made use of a 2D model for the Landau quasiparticles of the γ band (*23*), the parameters of which are all constrained by other observations on Sr_2RuO_4 (*25*). This model (*22*) yields a contribution $F_\gamma(T,\varepsilon_{xx},\varepsilon_{yy},\varepsilon_{zz})$ to the electronic free energy as function of temperature and the three uniaxial strain values. The crucial ingredient of the theory is a symmetry-adapted deformation potential $\alpha \sim t^{-1}\partial t/\partial\varepsilon$ with tight-binding hopping parameters t. (Notice that although we mostly refer to the uniaxial strain along the x axis as ε, we briefly write out the axis labels explicitly to account for the correct Poisson effects that enter any measurement of the Young's modulus.) The total free energy is then given as $F = F_0 + F_\gamma$, where we determined F_0 such that our model reproduces the correct elastic constants for the unstrained samples at the reference temperature of $T = 4$ K. We included the applied uniaxial stress using $F \rightarrow F - \varepsilon_{xx}\sigma_{xx}$, which yields the equation of state $\sigma_{xx} = \partial F/\partial\varepsilon_{xx}$. In linear elasticity, strain orthogonal to the applied stress is accounted for by Poisson ratios, such as $\varepsilon_{\kappa\kappa} = -\nu_{x\kappa}\varepsilon_{xx}$ with $\kappa = y, z$. For the rather large stress values applied here and given the subtle behavior near the Lifshitz transition, we must, however, allow for nonlinear relations $\varepsilon_{\kappa\kappa} = \varepsilon_{\kappa\kappa}(\varepsilon_{xx})$. Those follow from $\partial F/\partial\varepsilon_{yy} = \partial F/\partial\varepsilon_{zz} = 0$. This finally yields the differential Young's modulus

$$E_x = \frac{d\sigma_{xx}}{d\varepsilon_{xx}} = C_{11} + C_{12}\frac{\partial\varepsilon_{yy}(\varepsilon_{xx})}{\partial\varepsilon_{xx}} + C_{13}\frac{\partial\varepsilon_{zz}(\varepsilon_{xx})}{\partial\varepsilon_{xx}} \quad (1)$$

with the usual definition of the elastic tensor. The elastic constants $C_{ij} = C_{0,ij} - \frac{2}{k_B T}\int \frac{d^3k}{(2\pi)^3} f_{\mathbf{k}}(1 - f_{\mathbf{k}})\frac{\partial E_{\mathbf{k}}}{\partial\varepsilon_{ii}}\frac{\partial E_{\mathbf{k}}}{\partial\varepsilon_{jj}}$ consist of a background contribution $C_{0,ij}$ and the part due to the γ band with strain-dependent $E_{\mathbf{k}}$ and Fermi function $f_{\mathbf{k}}$. Within linear elasticity, the derivatives are the strain-independent Poisson ratios and $1/E_x$ is the $i = j = 1$ element of the inverse of the elastic tensor C_{ij} in Voigt notation. We show in (*22*) that, near the Lifshitz point, $-\varepsilon_{yy}/\varepsilon_{xx}$ and the differential Poisson ratio $-\partial\varepsilon_{yy}/\partial\varepsilon_{xx}$ differ and have a pronounced strain dependence (fig. S6). Once these nonlinear Poisson's ratios were included, our model made the predictions shown in Fig. 2, A and B. The agreement with the experimental data shown in Fig. 1 is very good (see also fig. S7). The temperature dependence of the dip in the Young's modulus and the relationship between the Young's modulus and entropy at 4 K are reproduced well and provide evidence that the model correctly captures the key physics of the observations.

Qualitatively, the softening at the Van Hove point is a consequence of the fact that Young's modulus is the sum of a presumed weakly temperature- and strain-dependent background contribution, $E_x^{(0)}$, and a singular conduction electron contribution from the γ band

$$E_x \cong E_x^{(0)} - A\log\frac{1}{(T/T_0)^2 + (\varepsilon_{xx} - \varepsilon_{\mathrm{VHS}})^2} \quad (2)$$

where A is a positive constant and T_0 is a temperature such that $k_B T_0$ is of the order of the electronic bandwidth. The logarithmic T and strain dependence stems from the fact that the electronic contribution to the elastic constants is proportional to the density of states [see the $f_{\mathbf{k}}(1-f_{\mathbf{k}})$ term in the γ-band contribution to C_{ij}], which diverges logarithmically at a VHS in 2D. All electronic contributions to the C_{ij} and the differential Poisson's ratios also show this singular, logarithmic dependence. The sign of the coefficient $A > 0$ reflects that the γ band will always cause a softening of the important diagonal elements C_{ii} of the elastic tensor. The magnitude of A is determined by a combination of band renormalization factors and the deformation potential: $A \propto \alpha^2$.

A further prediction of the model is that the temperature dependence of the Young's modulus at the Lifshitz pressure has negative curvature, and this is indeed observed in our measurements (Fig. 2C). This negative curvature, in addition to the strong link between elastic and electronic degrees of freedom that our data have established, imply that the Young's modulus is related to an electronic susceptibility. The logarithmic softening, however, cannot continue down to the lowest temperatures. This interruption might be a consequence of a first-order structural transition, which was initially proposed by Lifshitz, or the formation of some electronic order that prevents a mechanical instability. One way in which the logarithm will be cut off in a quasi-2D material is by coherent 3D effects caused by interlayer hopping. However, in the highly 2D γ band of Sr_2RuO_4, the scale for such processes is <3 K (*8*). Whether the onset of superconductivity at $T_c = 3.5$ K is related to the mechanical stability of Sr_2RuO_4 is therefore an exciting open question.

Interplay with superconductivity

In this context, it was interesting to extend our Young's modulus measurements to the superconducting state in which a small gap is opened at the Fermi energy. As shown in Fig. 3, A and B, the strong normal-state softening of the lattice is indeed cut off by the onset of the superconductivity, with the lattice hardening again slightly below the superconducting T_c, an effect that is most pronounced at the Van Hove strain. By contrast, if we suppress the superconductivity by applying a 2 T out-of-plane magnetic field, then the softening continues down to the lowest temperatures used in the measurement.

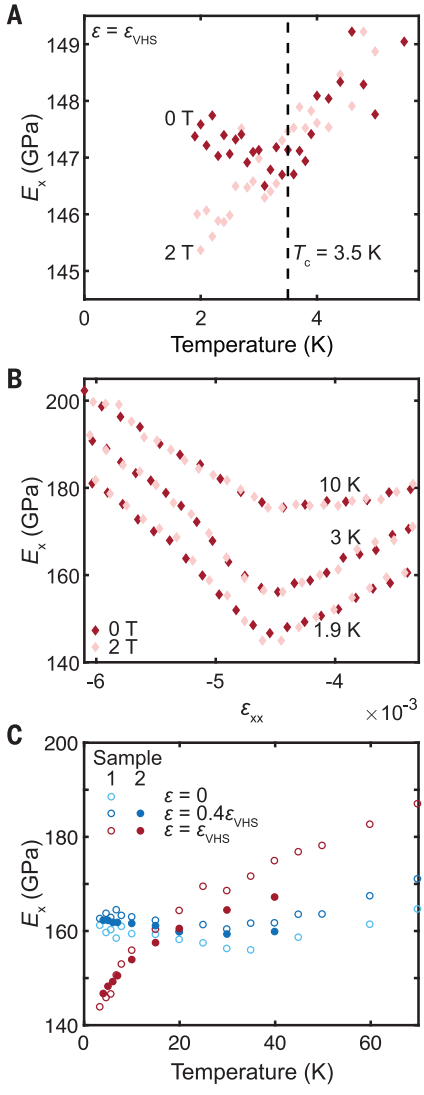


Fig. 3. Tracking E_x at key strains to lower and higher temperatures. (**A**) E_x at $\varepsilon_{\mathrm{VHS}}$ continues to soften with decreasing temperature as long as superconductivity is suppressed with a 2 T magnetic field (pink diamonds). In the presence of superconductivity, E_x hardens instead (0 T, red diamonds). The averaging window is a stress range of 0.05 GPa; the value of T_c is taken from (*19*). (**B**) E_x versus ε_{xx} at selected temperatures with and without an applied magnetic field (red, 0 T; pink, 2 T). The hardening is visible in the data at 1.9 K. The curves at 3 and 10 K have been offset by, respectively, 10 and 20 GPa for clarity. Data in (A) and (B) are from sample 3. (**C**) E_x versus temperature at $\varepsilon_{\mathrm{VHS}}$ (red), $0.4\varepsilon_{\mathrm{VHS}}$ (dark blue), and zero strain (light blue) over a wide range of temperatures. Open circles, sample 1; filled circles, sample 2. Data at $\varepsilon_{\mathrm{VHS}}$ for sample 2 are replotted from Fig. 2C. The averaging window is a strain range of 4.05×10^{-4} for sample 1 and 3×10^{-4} for sample 2.

A final aspect of our measurements is shown in Fig. 3C, in which we present E_x data over a wider range of temperatures, at zero strain, an intermediate strain, and the Van Hove strain. At zero strain, there is a broad minimum in the Young's modulus at $T \sim 40$ K, reflecting the fact that, unusually, Sr_2RuO_4 softens along the [100] crystalline direction as the temperature is decreased from room temperature (*26*). We show in fig. S8 that this feature is reproduced in our model, demonstrating that it too is a consequence of the conduction electrons in the γ sheet.

Discussion and outlook

Although the model that we have used to analyze our Sr_2RuO_4 data is specific to the Lifshitz transition in 2D, it can be viewed from a more general perspective that highlights the significance of low-temperature entanglement between electronic and elastic degrees of freedom and emphasizes the close connection between elastic response and entropy. Consider a quantum critical point (QCP) that can be crossed by varying some combination of elements of the strain tensor. Under the assumption that hyperscaling holds near any such a strain-tuned QCP, for the singular contribution to the entropy

$$S(T,\varepsilon) = T^{d/z}\phi\left(\frac{\varepsilon - \varepsilon_c}{T^{1/\nu z}}\right) \quad (3)$$

where z and ν are the dynamical and correlation length exponents, respectively, d is the dimension of space, ϕ is a universal scaling function, and ε_c is the critical strain. The most notable consequence of Eq. 3 follows as one integrates the entropy with respect to temperature to obtain the free energy and then determine the elastic constant. Right at the QCP, where $|\varepsilon - \varepsilon_c|^{\nu z} \ll T$, it follows that

$$C = C_0 - \frac{\nu z}{\nu(d+z) - 2}\phi''(0)T^{\frac{\nu(d+z)-2}{\nu z}} \quad (4)$$

where ϕ'' is the second derivative of ϕ. Here, C_0 is a temperature-independent background contribution to the elastic constant that enters as an integration constant for the free energy. Because the entropy is maximal at the QCP, $\phi'' < 0$. If $\nu(d+z) > 2$, then the universal temperature correction is small and positive and the system is mechanically stable. Conversely, the system must undergo an instability, defined by a vanishing elastic constant at a nonzero temperature, if the quantum Harris criterion $\nu(d+z) < 2$ (*27*, *28*) is fulfilled. Then, the above scaling theory of a "naked" QCP ceases to be valid; the system either crosses over to a new critical regime where strain becomes a genuine dynamical quantum critical mode or undergoes a phase transition to another state of matter. As discussed in (*22*) and figs. S9 and S10, the Lifshitz point herein corresponds to $d = z = \nu^{-1} = 2$, placing us at the boundary of the

quantum Harris criterion, resulting in a logarithmic temperature dependence, and leading to a mechanical instability. If one takes the limit $\nu(d+z) \rightarrow 2$, appropriate for the 2D Lifshitz transition, then the exponent in the temperature-dependent term in Eq. 4 does not simply vanish because the prefactor $\frac{1}{\nu(d+z)-2}$ diverges at the same time. Instead, one recovers the logarithmic behavior of the Lifshitz transition.

Returning to the discussion of the experimental findings, the results presented herein constitute conclusive observation of a lattice softening driven by conduction electrons, the possibility of which was foreseen by Lifshitz >60 years ago (*4*). However, he considered hydrostatic pressure and transitions in materials with 3D electronic structure, in which case the relative change in bulk modulus would be ~10^{-4}, three orders of magnitude smaller than what we observed [(*22*) and figs. S11 and S12]. Partly for this reason, previous searches using hydrostatic pressure were unable to unambiguously resolve the predicted effect (*29*–*33*). Why, then, does it give such a prominent experimental signature in Sr_2RuO_4? First, we have worked with uniaxial rather than hydrostatic pressure. Also, Sr_2RuO_4 is an extremely clean material for which the relevant band is strongly 2D, preventing the logarithmic term in Eq. 2 from being washed out by 3D effects or disorder broadening. It is tempting on first inspection to assume that this logarithm makes the dominant contribution to the size of our signal, but for measurements performed at a few degrees kelvin, the size of the prefactor A actually plays the crucial role. It is hugely enhanced over Lifshitz's original expectation for three reasons: (i) the correlation-induced γ band renormalization, (ii) the nonlinear Poisson's ratio effect contribution to Young's modulus, and (iii) the value of α in the deformation potential. In our model, the experimentally observed Van Hove strain yields $\alpha = 7.6$ and enters A as α^2. To investigate further, we performed first-principles calculations, which emphasized that the γ band of Sr_2RuO_4 is based on Ru-O-Ru processes involving two d–p orbital hops, yielding $\alpha = 8$ [(*22*) and fig. S13].

Our findings also provide perspectives on the nature and consequences of the entanglement between elastic and electronic degrees of freedom in metallic solids. To what extent might they be a driver for superconductivity as a route to avoid divergent lattice softening? Do the physics directly following from conduction electron density of states play a bigger than previously appreciated role in effects in heavy fermion physics such as the Kondo volume collapse (*34*) and lattice softening associated with magnetism (*35*) and metamagnetism (*36*)? Although these remain open questions, our observations provide strong and concrete evidence of relevance to the "chicken and egg" problem discussed in the introduction: Conduction band physics can drive unexpectedly large structural effects, and conduction electrons are not always slaves to the lattice.

REFERENCES AND NOTES

1. E. Fradkin, S. A. Kivelson, M. J. Lawler, J. P. Eisenstein, A. P. Mackenzie, *Annu. Rev. Condens. Matter Phys.* **1**, 153–178 (2010).
2. R. Fernandes, A. Chubukov, J. Schmalian, *Nat. Phys.* **10**, 97–104 (2014).
3. J.-H. Chu, H.-H. Kuo, J. G. Analytis, I. R. Fisher, *Science* **337**, 710–712 (2012).
4. I. M. Lifshitz, *Sov. Phys. JETP* **11**, 1130–1135 (1960).
5. M. E. Barber, A. S. Gibbs, Y. Maeno, A. P. Mackenzie, C. W. Hicks, *Phys. Rev. Lett.* **120**, 076602 (2018).
6. V. Sunko *et al*., *NPJ Quantum Mater.* **4**, 46 (2019).
7. L. Van Hove, *Phys. Rev.* **89**, 1189–1193 (1953).
8. A. P. Mackenzie, Y. Maeno, *Rev. Mod. Phys.* **75**, 657–712 (2003).
9. Y. Maeno, S. Kittaka, T. Nomura, S. Yonezawa, K. Ishida, *J. Phys. Soc. Jpn.* **81**, 011009 (2012).
10. C. Kallin, *Rep. Prog. Phys.* **75**, 042501 (2012).
11. Y. Liu, Z.-Q. Mao, *Physica C* **514**, 339–353 (2015).
12. A. P. Mackenzie, T. Scaffidi, C. W. Hicks, Y. Maeno, *NPJ Quantum Mater.* **2**, 40 (2017).
13. Y. Maeno *et al*., *J. Phys. Soc. Jpn.* **66**, 1405–1408 (1997).
14. C. Bergemann, A. P. Mackenzie, S. R. Julian, D. Forsythe, E. Ohmichi, *Adv. Phys.* **52**, 639–725 (2003).
15. J. S. Bobowski *et al*., *Condens. Matter* **4**, 6 (2019).
16. K. M. Shen *et al*., *Phys. Rev. Lett.* **99**, 187001 (2007).
17. A. Steppke *et al*., *Science* **355**, eaaf9398 (2017).
18. M. E. Barber *et al*., *Phys. Rev. B* **100**, 245139 (2019).
19. Y.-S. Li *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020492118 (2021).
20. M. E. Barber, A. Steppke, A. P. Mackenzie, C. W. Hicks, *Rev. Sci. Instrum.* **90**, 023904 (2019).
21. M. Ikhlas *et al*., *Appl. Phys. Lett.* **117**, 233502 (2020).
22. See the supplementary materials.
23. Y.-S. Li *et al*., *Nature* **607**, 276–280 (2022).
24. A. P. Mackenzie *et al*., *J. Phys. Soc. Jpn.* **67**, 385–388 (1998).
25. B. Burganov *et al*., *Phys. Rev. Lett.* **116**, 197003 (2016).
26. N. Okuda, T. Suzuki, Z. Mao, Y. Maeno, T. Fujita, *J. Phys. Soc. Jpn.* **71**, 1134–1139 (2002).
27. M. Zacharias, A. Rosch, M. Garst, *Eur. Phys. J. Spec. Top.* **224**, 1021–1040 (2015).
28. P. Chandra, P. Coleman, M. A. Continentino, G. G. Lonzarich, *Phys. Rev. Res.* **2**, 043440 (2020).
29. L. Dubrovinsky *et al*., *Nature* **525**, 226–229 (2015).
30. G. Steinle-Neumann, L. Stixrude, R. E. Cohen, *Phys. Rev. B* **63**, 054103 (2001).
31. G. Aquilanti *et al*., *Phys. Rev. B* **76**, 144102 (2007).
32. B. K. Godwal, S. V. Raju, Z. Geballe, R. Jeanloz, *J. Phys. Conf. Ser.* **377**, 012033 (2012).
33. Ya. M. Blanter, M. I. Kaganov, A. V. Pantsulaya, A. A. Varlamov, *Phys. Rep.* **245**, 159–257 (1994).
34. J. W. Allen, R. M. Martin, *Phys. Rev. Lett.* **49**, 1106–1110 (1982).
35. H. Rosner *et al*., *Nat. Phys.* **2**, 469–472 (2006).
36. F. Weickert, M. Brando, F. Steglich, P. Gegenwart, M. Garst, *Phys. Rev. B* **81**, 134438 (2010).
37. Data for: H. M. L. Noad *et al*., Giant lattice softening at a Lifshitz transition in Sr_2RuO_4, Edmond (2023); https://doi.org/10.17617/3.UOVU6O.

ACKNOWLEDGMENTS

Funding: This work was supported by the Max Planck Society; Research in Dresden benefits from the environment provided by the DFG Cluster of Excellence (ct.qmat EXC 2147, Project 390858940 to A.P.M.); the German Research Foundation (TRR 288-422213477 ELASTO-Q-MAT, Project A10 to H.M.L.N., A.P.M., and C.W.H.; TRR 288-422213477 ELASTO-Q-MAT, Project A11 to M.G.; and TRR 288-422213477 ELASTO-Q-MAT, Project B01 to J.S. and V.S.); the Alexander von Humboldt Foundation (Research Fellowship for Postdoctoral Researchers to H.M.L.N.); received funding of the Klaus Tschira Boost Fund, a joint initiative of the German Scholars Organization and the Klaus Tschira Foundation (to H.M.L.N.); the Japan Society for the Promotion of Science (Overseas Research Fellowship to K.I. and KAKENHI Grants-in-Aid for Scientific Research grants 17H06136, 18K04715, 21H01033, and 22K19093 to N.K.; and Core-to-Core Program grant JPJSCCA20170002 to N.K.); the Japan Science and Technology Agency (JST)–Mirai Program (grant no. JPMJMI18A3 to N.K.); and the National Research Foundation of Korea (grants 2021R1C1C1007017 and 2022M3H4A1A04074153 to B.K.). **Author contributions:** Conceptualization: H.M.L.N., J.S., C.W.H.; Formal analysis: H.M.L.N., M.G., J.S., C.W.H.; Funding acquisition: M.G., J.S., A.P.M., C.W.H.; Investigation: H.M.L.N., K.I., Y.-S.L., E.G., V.S., B.K., I.I.M., M.G., J.S.; Methodology: H.M.L.N., E.G., V.S., J.S., C.W.H.; Project administration: H.M.L.N., A.P.M.; Resources: N.K., D.A.S.; Supervision: M.N., I.I.M., J.S., A.P.M., C.W.H.; Visualization: H.M.L.N.; Writing – original draft: H.M.L.N., M.G., J.S., B.K., I.I.M., A.P.M., C.W.H.; Writing – review and editing: H.M.L.N., K.I., Y.-S.L., N.K., M.N., B.K., I.I.M., M.G., J.S., A.P.M., C.W.H. **Competing interests:** C.W.H. has 31% ownership of Razorbill Instruments, a company that markets uniaxial pressure cells. C.W.H. is a coapplicant on patent application GB1711384.6, which relates to the methodology in this work. The remaining authors declare no competing interests. **Data and materials availability:** Data in the main text and supplementary materials are available at Edmond, the open research data repository of the Max Planck Society (*37*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf3348
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*38*–*51*)

Submitted 19 October 2022; accepted 16 September 2023
10.1126/science.adf3348

DIFFERENTIATION

Unbiased transcription factor CRISPR screen identifies ZNF800 as master repressor of enteroendocrine differentiation

Lin Lin[1,2,3]*, Jeff DeMartino[2,3], Daisong Wang[1,2], Gijs J. F. van Son[2,3], Reinier van der Linden[1,2], Harry Begthel[1,2], Jeroen Korving[1,2], Amanda Andersson-Rolf[1,2], Stieneke van den Brink[1,2], Carmen Lopez-Iglesias[4], Willine J. van de Wetering[4], Aleksandra Balwierz[3], Thanasis Margaritis[3], Marc van de Wetering[2,3], Peter J. Peters[4], Jarno Drost[2,3], Johan H. van Es[1,2], Hans Clevers[1,2,3]*†

Enteroendocrine cells (EECs) are hormone-producing cells residing in the epithelium of stomach, small intestine (SI), and colon. EECs regulate aspects of metabolic activity, including insulin levels, satiety, gastrointestinal secretion, and motility. The generation of different EEC lineages is not completely understood. In this work, we report a CRISPR knockout screen of the entire repertoire of transcription factors (TFs) in adult human SI organoids to identify dominant TFs controlling EEC differentiation. We discovered ZNF800 as a master repressor for endocrine lineage commitment, which particularly restricts enterochromaffin cell differentiation by directly controlling an endocrine TF network centered on PAX4. Thus, organoid models allow unbiased functional CRISPR screens for genes that program cell fate.

Enteroendocrine cells (EECs) are specialized epithelial cells of the intestinal tract that, like all other epithelial cell lineages, derive from regionally specified Lgr5$^+$ intestinal stem cells (ISCs) (*1*–*5*). Balanced differentiation of EEC lineages from ISCs is governed by a network of transcription factors (TFs) (*6*–*12*). RNA sequencing and organoid technology have illuminated the temporal hierarchy of gene expression profiles during EEC development (*13*, *14*). By using a time-resolved reporter allele of *Neurog3* (master regulator of endocrine development) (*15*, *16*), a real-time and lineage-specific map of mouse EEC differentiation in vivo was constructed at single cell–level resolution (*12*). Overexpression of *NEUROG3* can generate endocrine cells for functional characterization in human pancreatic duct cells (*17*). Leveraging human gut organoids, this strategy enabled profiling of EEC subtypes along the proximal-distal gastrointestinal axis (*18*). TFs known to specify EEC subtypes in mice were further examined in human organoids, revealing a predominantly conserved regulatory mechanism downstream of *NEUROG3* activation. However, discrepancies between mouse and human data were also observed. The regulatory mechanisms upstream of *NEUROG3* and the endogenous repressive factors that control *NEUROG3* expression at the adult stage remain largely unknown. In this work, we established an organoid-based platform in combination with high-throughput genetic screening for the unbiased discovery of TFs that govern human EEC lineage commitment.

Results

TFome-wide CRISPR knockout screen in human SI organoids

Optimized human SI organoids exhibit the spontaneous generation of all major cell lineages, including EEC subtypes, without TF overexpression (*19*). For the current study, we aimed to establish a CRISPR screening platform spanning the entire repertoire of transcription factors (TFome) in human intestinal organoids to allow for unbiased and systematic discovery of TFs that govern cell lineage commitment from adult LGR5$^+$ intestinal stem cells. We used a CRISPR knockout library comprising 7210 single-guide RNAs (sgRNAs) targeting 1800 human TFs, alongside 100 negative controls (*20*–*22*). In the human ileum organoid line used, EECs were marked by CHGA-IRES-iRFP670, goblet cells by MUC2-mNeonGreen, and Paneth cells by DEFA5-IRES-DsRed (fig. S1A). We performed lentiviral-mediated library integration in expanding SI organoids (Fig. 1A). Library coverage was assessed by sequencing the integrated sgRNAs, comparing the organoids after transduction with the plasmid pool.

Fate separation of CHGA$^+$ EECs and MUC2$^+$ cells occurs early in secretory progenitors, whereas DEFA5$^+$ Paneth cells emerge later from MUC2$^+$ progenitors (*19*), illustrating sequential fate decisions of ISCs (*23*, *24*). We aimed to track the early CHGA$^+$/MUC2$^+$ binary fate decision. After two-step organoid differentiation (*19*), we analyzed bulk populations as well as sorted EECs (CHGA$^+$), goblet cells (MUC[illegible]) and the CHGA$^-$/MUC2$^-$/DEFA5$^-$ population (fig. S1B). 0 to ~0.7% sgRNAs yielded zero reads in bulk populations, indicating robust coverage (fig. S2A). Normalized read-count distribution and Pearson correlations demonstrated high concordance between biological replicates (fig. S2, B and C). sgRNAs that target essential ISC genes [*KLF5* (*25*) and *TCF7L2*/*TCF4* (*26*–*29*)] were depleted, along with "generic" cell fitness genes (*POLR2L*, *MYC*, and *RAD51*) at both expansion and differentiation culturing stages (β score < −1; FDR < 0.05 versus plasmid library) (fig. S2D and table S1). Furthermore, sgRNAs targeting *IRF2*, an interferon pathway TF, were highly enriched upon differentiation, consistent with existing literature (*30*, *31*).

sgRNAs were then assessed in CHGA$^+$ EECs and MUC2$^+$ goblet cells versus the triple reporter–negative cell population (Fig. 1B and table S2). As expected, *NEUROG3*, *SOX4*, and *INSM1* sgRNAs were depleted in EECs (*8*, *15*, *16*, *32*). Additionally, *NFIC*, *TEF*, and *ZHX2* appeared essential for EEC commitment. For each of these, clonal knockout organoids revealed a mild yet significant reduction of CHGA$^+$ EECs (Fig. 1, B and C, and fig. S3, A to C). Conversely, *GFI1* sgRNAs were enriched in EECs (Fig. 1, B and C). Mouse *Gfi1* is crucial for goblet cell differentiation by suppression of *Neurog3*-driven EEC cell fate (*23*, *33*). Indeed, *GFI1* knockout in human organoids abrogated goblet cell formation while increasing EEC numbers (fig. S3, A and C, and fig. S14, C and D).

ZNF800 represses EEC differentiation

The strongest repressor of EEC differentiation in the screen, *ZNF800*, is a C_2H_2 zinc-finger TF of unknown function (Fig. 1, B and C) (*34*). Human *ZNF800* is broadly expressed, including in the SI and colon epithelium (fig. S4, A and B). ZNF800$^{-/-}$ organoids contained increased EECs and strongly reduced goblet and Paneth cell numbers (Fig. 1, D and E, and fig. S4, C and G). Transmission electron microscopy confirmed the absence of goblet and Paneth cells with the characteristic apical secretory vesicles and the increase of EECs with the basolateral secretory granules (Fig. 1F). Similarly, ZNF800 knockout in colon organoid lines from two different donors also resulted in a significant increase in EECs (fig. S4, D to G).

We next performed single-cell RNA sequencing (scRNA-seq) to study the phenotype of ZNF800$^{-/-}$ organoids. For equal representation, CHGA$^+$ cells were sorted from wild type (WT) and ZNF800$^{-/-}$ organoids and pooled with CHGA$^-$ cells from the same lines in a 1:4 ratio (Fig. 2A and fig. S5A). Major intestinal cell types were identified by graph-based clustering analysis (Fig. 2B and fig. S5B). We observed the expected reduction of goblet cells

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, Utrecht, Netherlands. [2]Oncode Institute, Utrecht, Netherlands. [3]Princess Maxima Center for Pediatric Oncology, Utrecht, Netherlands. [4]The Maastricht Multimodal Molecular Imaging Institute, Maastricht University, Maastricht, Netherlands.
*Corresponding author. Email: l.lin@hubrecht.eu (L.L.); h.clevers@hubrecht.eu (H.C.)
†Current address: Pharma Research and Early Development of F. Hoffmann–La Roche Ltd., Basel, Switzerland.

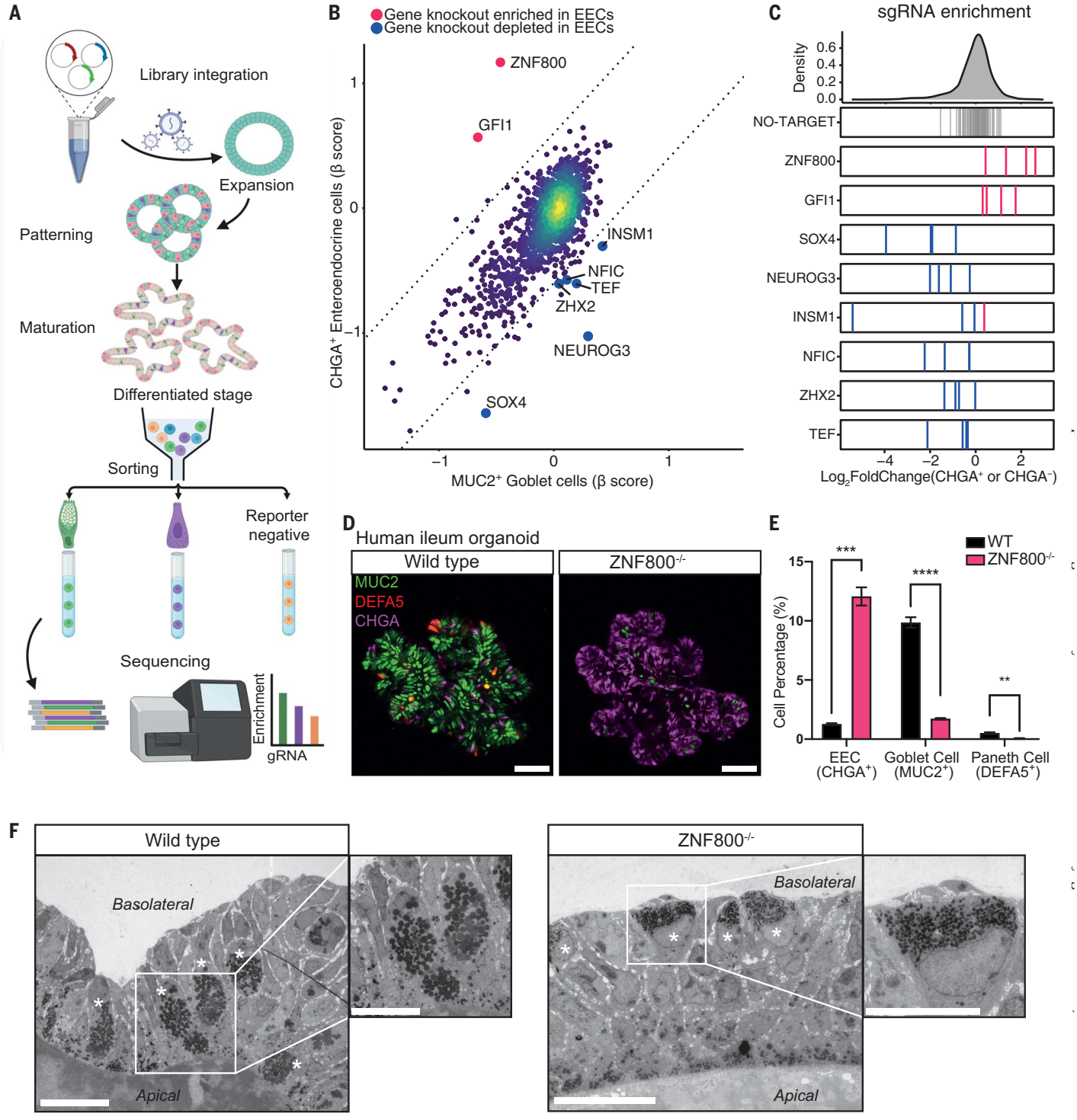


Fig. 1. TFome-wide CRISPR screen for endocrine differentiation in human SI organoids. (**A**) Schematic of TFome CRISPR screen. (**B**) Scatter plot of enrichment β score of each TF gene in CHGA$^+$ EECs and MUC2$^+$ goblet cells versus the triple reporter–negative cell population. (**C**) Individual sgRNA enrichment for genes of interest presented by $\log_2$-fold changes in CHGA$^+$ EECs versus triple reporter–negative cell population. Density plot (top) represents the distribution of nontargeting sgRNAs. (**D**) Representative confocal images of WT and ZNF800$^{-/-}$ human SI organoids. Representative marker genes for EECs (CHGA, magenta), goblet cells (MUC2, green), and Paneth cells (DEFA5, red) are highlighted by fluorescent reporters. Scale bars, 100 μm. (**E**) Proportion of different differentiated cell types as determined by FACS analysis of the respective reporters. Data are shown as mean ± SEM. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$ by multiple *t* tests with two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli with $Q = 5\%$ and $n = 3$. (**F**) Transmission electron microscopy images of WT and ZNF800$^{-/-}$ human SI organoids. Goblet and Paneth cells are indicated with asterisks in WT organoids, and EECs are indicated with asterisks in ZNF800$^{-/-}$ organoids. Scale bars, 20 μm and 10 μm (zoom in).

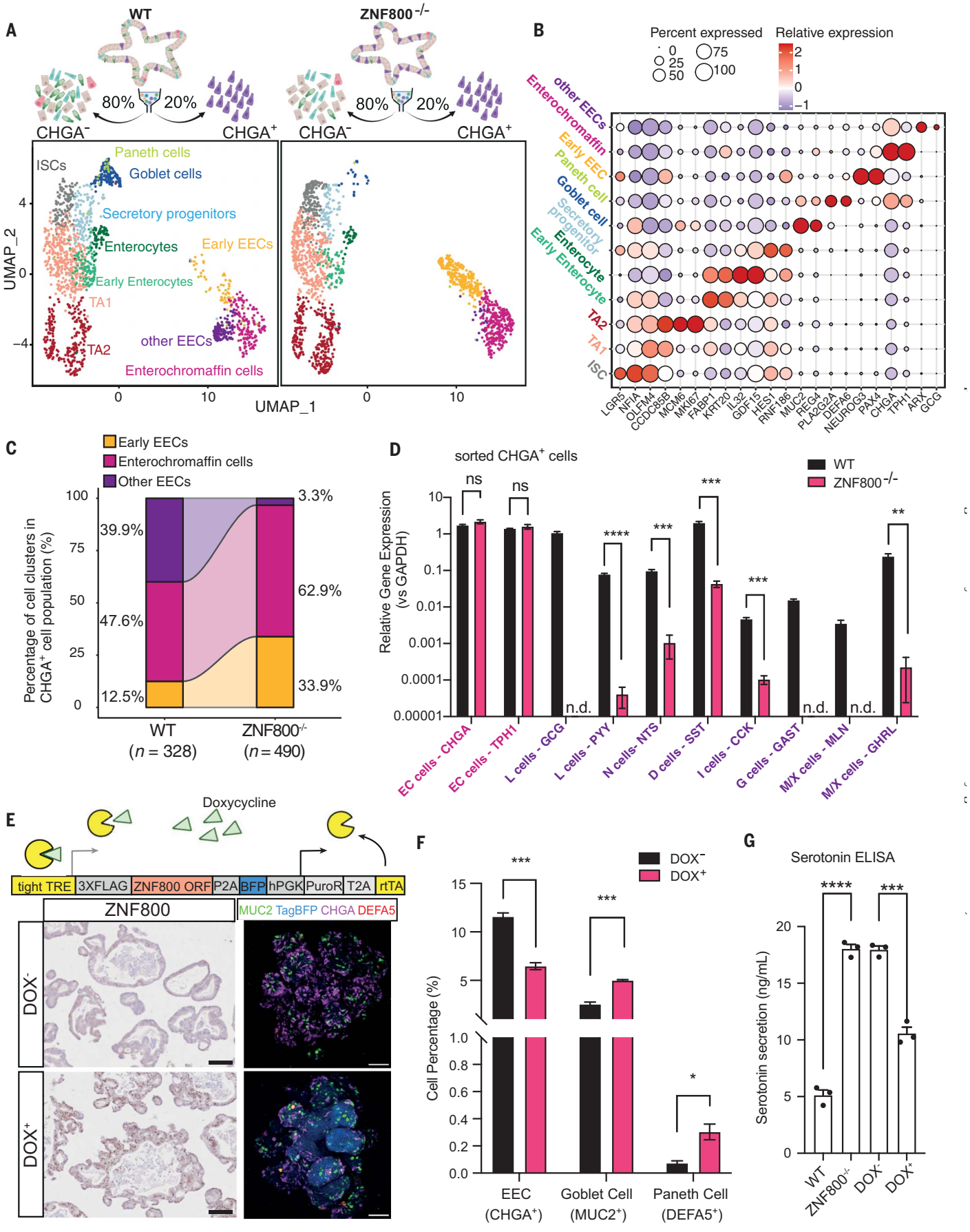

A
WT
ZNF800 -/-
80%
20%
CHGA⁻
CHGA⁺
ISCs
Paneth cells
Goblet cells
Secretory progenitors
Enterocytes
Early EECs
Early Enterocytes
TA1
TA2
other EECs
Enterochromaffin cells
UMAP_2
UMAP_1
B
Percent expressed
Relative expression
other EECs
Enterochromaffin
Early EEC
Paneth cell
Goblet cell
Secretory progenitor
Enterocyte
Early Enterocyte
TA2
TA1
ISC
LGR5
NFIA
OLFM4
CCDC85B
MCM6
MKI67
FABP1
KRT20
IL32
GDF15
HES1
RNF186
MUC2
REG4
PLA2G2A
DEFA6
NEUROG3
PAX4
CHGA
TPH1
ARX
GCG
C
Early EECs
Enterochromaffin cells
Other EECs
Percentage of cell clusters in CHGA⁺ cell population (%)
39.9%
47.6%
12.5%
3.3%
62.9%
33.9%
WT (n = 328)
ZNF800-/- (n = 490)
D
sorted CHGA⁺ cells
WT
ZNF800 -/-
Relative Gene Expression (vs GAPDH)
ns
****
***
**
n.d.
EC cells - CHGA
EC cells - TPH1
L cells - GCG
L cells - PYY
N cells- NTS
D cells - SST
I cells - CCK
G cells - GAST
M/X cells - MLN
M/X cells - GHRL
E
Doxycycline
tight TRE
3XFLAG
ZNF800 ORF
P2A
BFP
hPGK
PuroR
T2A
rtTA
ZNF800
MUC2 TagBFP CHGA DEFA5
DOX-
DOX+
F
DOX⁻
DOX⁺
Cell Percentage (%)
EEC (CHGA⁺)
Goblet Cell (MUC2⁺)
Paneth Cell (DEFA5⁺)
G
Serotonin ELISA
Serotonin secretion (ng/mL)
WT
ZNF800-/-
DOX-
DOX+

Fig. 2. Phenotypic characterization of ZNF800$^{-/-}$ organoids. (**A**) Uniform manifold approximation and projection (UMAP) of WT and ZNF800$^{-/-}$ organoids from scRNA-seq. Descriptive cluster labels are shown. TA1, transit-amplifying cell stage 1; TA2, transit-amplifying cell stage 2. (**B**) Dot plot showing the relative expression and the percentage of cells expressing selected markers across scRNA-seq clusters. Two representative markers for each cluster are plotted. (**C**) Stacked bar plot showing the EEC cell populations in WT and ZNF800$^{-/-}$ organoids. (**D**) Reverse transcription (RT)–qPCR quantification of various EEC markers in CHGA^{+} cell populations of WT and ZNF800$^{-/-}$ organoids. (**E**) (Top) Schematic of dox-inducible overexpression construct of ZNF800. (Bottom left) Representative immunohistochemical staining images of ZNF800 in ZNF800$^{-/-}$ organoids with or without dox-induced ZNF800 expression. (Bottom right) Representative confocal images of organoids with fluorescent reporters. Scale bars, 100 μm. (**F**) Proportion of EECs and goblet and Paneth cells as determined by fluorescence-activated cell sorting (FACS) analysis of the respective reporters in ZNF800$^{-/-}$ organoids with or without dox-induced ZNF800 expression. (**G**) Enzyme-linked immunosorbent assay (ELISA) quantification of serotonin secretion of organoids in different conditions. Data in this figure are shown as mean ± SEM. N.d. not detected; ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 by multiple t tests with two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli with Q = 5% and n = 3.

in ZNF800$^{-/-}$ organoids (2.3%) compared with WT organoids (10.8%) (fig. S5C). Clonal formation efficiency of WT and ZNF800$^{-/-}$ SI and colon organoids revealed no significant differences (fig. S6, A and B). Cell-cycle progression in WT versus ZNF800$^{-/-}$ organoids using a Fucci cell-cycle reporter (*35*) showed no significant differences (fig. S6C). The combined results indicate that ZNF800 does not affect ISC homeostasis in organoids.

EECs were further subclustered based on markers of early EECs (*NEUROG3* and *PAX4*), enterochromaffin cells (ECs, *CHGA* and *TPH1*) and the other EEC subtypes (D cells, *SST*; L cells, *GCG* and *PYY*; M/X cells, *MLN* and *GHRL*; I cells, *CCK*; G cells, *GAST*; N cells, *NTS*) (Fig. 2B and fig. S7A). Notably, early EECs and ECs were increased in ZNF800$^{-/-}$ organoids, whereas other EEC subtypes were depleted (Fig. 2, C and D, and fig. S7A). The same phenotype was also observed in human colon organoids (fig. S8, A and B). Thus, *ZNF800* loss induces robust EEC differentiation at the expense of goblet cells while also driving an EC-biased trajectory.

We reexpressed WT ZNF800 protein in ZNF800$^{-/-}$ organoids by doxycycline (dox)–inducible overexpression (Fig. 2E and fig. S8C), which effectively rescued goblet and Paneth cell differentiation while repressing EEC lineage commitment (Fig. 2F). Notably, the EC-biased differentiation pattern was also reversed upon ZNF800 rescue, which resulted in decreased serotonin secretion, a hallmark hormone of ECs (Fig. 2G and fig. S8D).

ZNF800 represses the endocrine TF regulatory network

Differential gene expression (DGE) analysis in CHGA^{+} cell populations of WT and ZNF800$^{-/-}$ organoids highlighted distinct expression patterns of EEC-specific TFs (fig. S9, A and B). To unravel these gene regulatory mechanisms, we performed single-cell regulatory network inference and clustering (SCENIC) analysis on our scRNA-seq dataset (*36*), which identified 249 regulons activated across different cell clusters (table S4). Regulons controlling EEC commitment displayed higher activity in ZNF800$^{-/-}$ organoids, including *SOX4*, *NEUROD2*, and *PAX4* (Fig. 3A). Notably, *TEF* and *NFIC*, discovered in the TFome CRISPR screen as EEC regulators, also showed regulon activity in EEC lineages (fig. S9C), supporting the functional relevance of our hits.

Chromatin immunoprecipitation sequencing (ChIP-seq) with anti-ZNF800 and anti-FLAG antibodies revealed 11,565 consensus peaks in WT organoids and 7085 in rescued ZNF800$^{-/-}$ organoids (q value < 0.01) (fig. S10A). Most ZNF800-binding sites localized within ±5 kb of transcription start sites (TSS) (fig. S10B). Comparison between ZNF800 peak-proximal genes in WT organoids and in rescued ZNF800$^{-/-}$ organoids showed a clear overlap (Fig. 3B). Of the consensus 3625 ZNF800-binding genes, we captured 3085 (85%) in our scRNA-seq dataset (fig. S11A). DGE analysis (adjusted P value < 0.05) revealed that 461 (15%) of the ZNF800-bound genes were significantly upregulated upon ZNF800 knockout, whereas 59 (2%) were significantly down-regulated (fig. S11B and table S5), indicating that ZNF800 functions as a transcriptional repressor.

Gene ontology enrichment analysis of 870 genes (corresponding to the top 1000 anti-ZNF800 ChIP-seq peak loci) revealed enrichment of neural and endocrine-gland development pathways (fig. S11C and table S6). We then constructed an enrichment map for gene sets involved in gland development, endocrine system development, and pancreas development (Fig. 3C). In particular, *INSM1*, *NEUROG3*, and *PAX4* were found to be central TFs in each node with the highest ZNF800 binding activity. As predicted from SCENIC (Fig. 3A), ZNF800 bound *SOX4* and *NEUROD2* loci (Fig. 3D and fig. S12A). *EGR2* and *DLL3* were also prioritized as top hits by ZNF800 peak abundance; both are implicated in neurogenesis (*37*–*41*) and specifically expressed in endocrine cell lineages (fig. S12B). These findings were highly correlated with their respective gene expression profiles (fig. S10C and S12B). Notably, we also identified mild enrichment of ZNF800 binding activity on the *NEUROD1* locus (fig. S12A). Knockout of *ZNF800* resulted in increased gene expression of *NEUROD1* (fig. S11B). However, *NEUROD1* did not appear as a hit in our CRISPR screen (fig. S12C). *Neurod1* acts as a late-stage EEC TF in mice (*12*). In human EECs, *NEUROD1* was also found to mark late progenitors and mature EECs (*18*). We therefore generated NEUROD1$^{-/-}$ organoids, which exhibited no effect on EEC or goblet cell differentiation (fig. S12, D to F), whereas profiling of EEC subtypes demonstrated a mild decrease in ECs, G cells, and M/X cells and a mild increase in N cells (fig. S12G). These results confirmed that *NEUROD1* is a late-stage TF regulator, which does not affect the EEC or goblet binary cell–fate decision.

Most ZNF800-binding loci presented a bivalent (H3K4me3- and H3K27me3-containing) chromatin signature in the SI and colon (Fig. 3D and fig. S13A). Such marks are associated with stem-cell differentiation, allowing timely responsiveness to TF regulation (*42*, *43*). The reversibility of gene activation or suppression by poised ZNF800-binding loci was tested in organoids with inducible ZNF800 expression (fig. S13B). Sequential cycles of dox induction demonstrated that ZNF800 alone can trigger a dynamic yet reversible equilibrium between EECs and goblet cells.

GFI1 and ZNF800 function independently in repressing EEC differentiation

Given that both ZNF800 and GFI1 were discovered in our screen as repressors of EEC differentiation, we sought to understand their regulatory interactions by generating a double knockout of *GFI1* and *ZNF800* (GFI1$^{-/-}$;ZNF800$^{-/-}$) in human SI organoids (fig. S14A). Further abrogation of goblet and Paneth cells was observed in GFI1$^{-/-}$;ZNF800$^{-/-}$ organoids (fig. S14, C and D), mirrored by further induction of EECs. The GFI1$^{-/-}$ single knockout did not lead to EC bias (fig. S14E). Instead, we observed a mild decrease of EC cells (*TPH1*) and a subtle increase in L cells (*PYY*). Notably, ZNF800 expression was not affected by *GFI1* loss (fig. S14, B and E). Because GFI1 is expressed in goblet and Paneth cells (*33*) and the knockout of *ZNF800* significantly depleted both cell types, we assayed GFI1 expression in goblet cells sorted from WT (12 ± 1.0%) and ZNF800$^{-/-}$ (3.4 ± 0.14%) organoids to avoid biased cell heterogeneity of the organoids, which revealed no differences in GFI1 expression (fig. S14F). Overall, our results indicated that ZNF800 and GFI1 function independently as repressors of EEC differentiation.

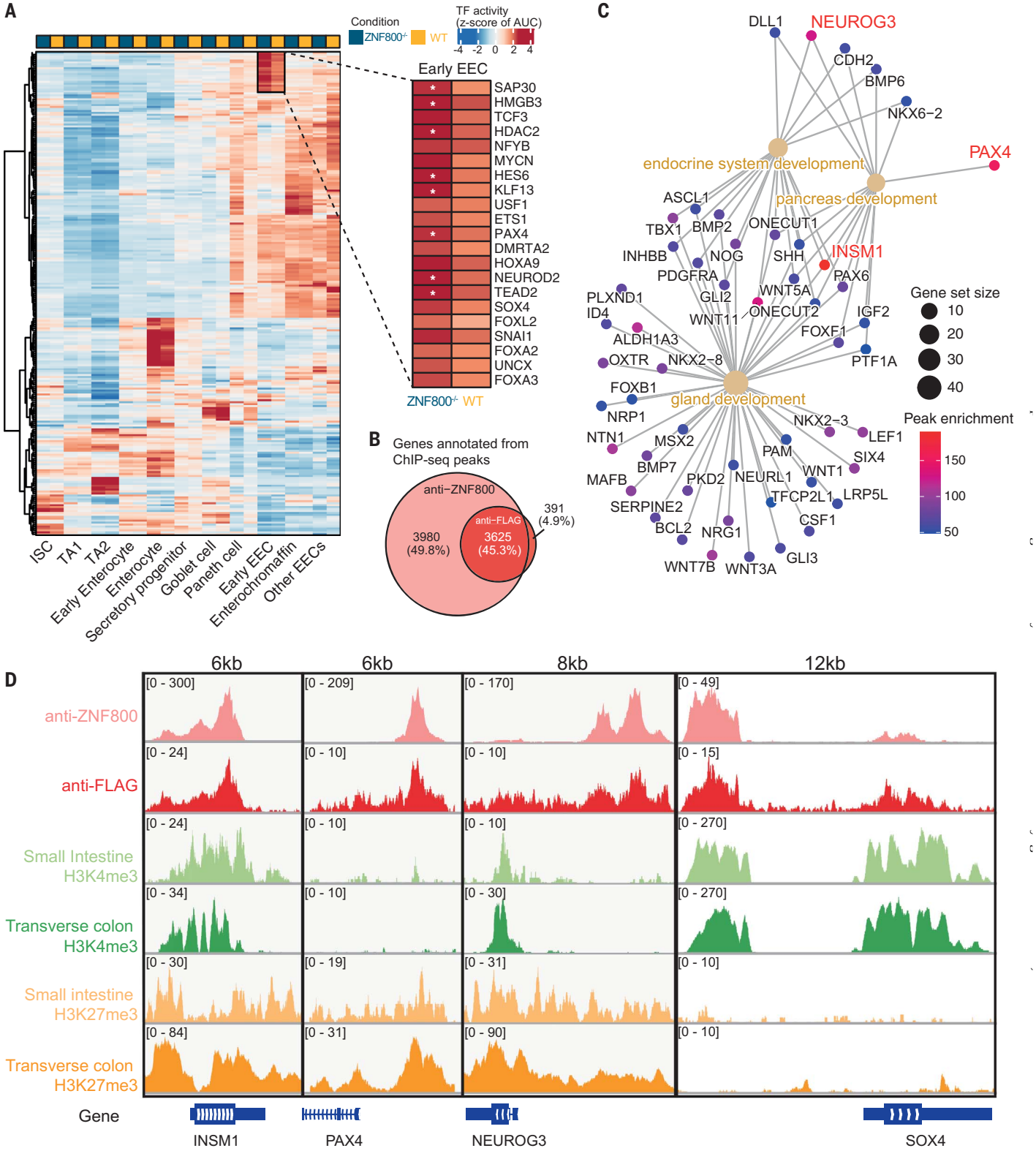


Fig. 3. ZNF800 binds to chromatin regions of TFs involved in endocrine differentiation. (**A**) Heatmap with unsupervised clustering of gene regulatory network activity in different cell clusters within WT and ZNF800$^{-/-}$ organoids and visualized as row *z*-scores of mean area-under-the-curve (AUC) values. Zoom-in plot highlighting the top TF regulons specifically activated in early EECs. Regulons with significant differences between WT and ZNF800$^{-/-}$ are indicated with asterisks, assessed by Wilcoxon rank sum test. (**B**) Venn diagram of overlapping genes annotated from ChIP-seq datasets generated by anti-ZNF800 antibody in WT organoids and anti-FLAG antibody in ZNF800$^{-/-}$ organoids with dox-induced ZNF800 expression. (**C**) An enrichment-network plot depicting the linkages of gene sets and three selected pathways by gene ontology analysis. (**D**) ChIP-seq tracks at the *INSM1*, *PAX4*, *NEUROG3*, and *SOX4* loci.

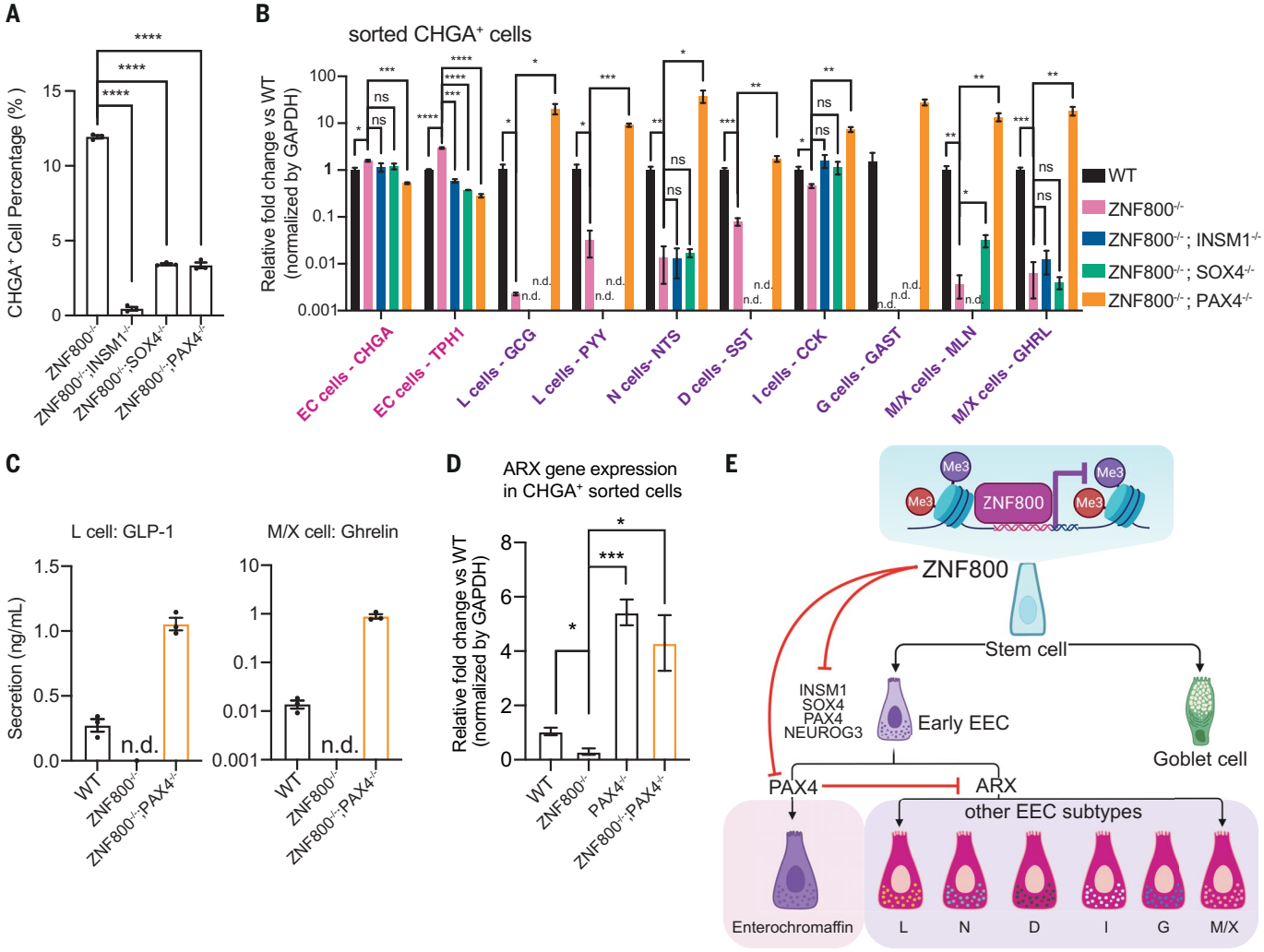


Fig. 4. PAX4 is responsible for EC-biased differentiation trajectory as a direct downstream target of ZNF800. (**A**) Proportions of EECs as determined by FACS analysis of organoids from different genotypes by CHGA reporter. (**B**) RT-qPCR quantification of various EEC markers in CHGA$^+$ cell populations of organoids from different genotypes. (**C**) ELISA quantification of GLP-1 and Ghrelin secretion of organoids from different genotypes. (**D**) RT-qPCR quantification of ARX expression in CHGA$^+$ cell populations of organoids from different conditions. (**E**) Schematic of ZNF800-driven repression model during endocrine cell differentiation. Data in this figure are shown as mean ± SEM. N.d. not detected; ns, not significant. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 by multiple t tests with two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli with Q = 5%, n = 3.

PAX4 drives EC differentiation directly downstream of ZNF800

A panel of genes was found to be up-regulated across multiple cell types upon ZNF800 knockout (table S7), including *COL4A2*, *TMEM178B*, and *BTBD11* (fig. S15A). ZNF800 binds to these genes, but not to the down-regulated genes such as *SULT1C2*, *HMGCS1*, and *MS4A8* (fig. S15B). Despite their modest up-regulation, the expression pattern and the annotated function of these genes did not directly relate to the EEC phenotype. Therefore, we focused on EEC TFs with stronger functional implications (fig. S9, A and B).

We focused on three direct ZNF800 target genes, *INSM1*, *SOX4*, and *PAX4* (Fig. 3D), by performing double knockouts (fig. S16A). EEC induction was effectively suppressed, but to different extents (Fig. 4A). Similar to the single depletion of *INSM1* (fig. S3C), *INSM1* knockout in ZNF800$^{-/-}$ organoids caused an essentially complete reversal of EEC induction, confirming its pivotal role in the TF cascade (*8*, *44*). Additional knockout of *SOX4* or *PAX4* in ZNF800$^{-/-}$ organoids both reduced EEC numbers.

Next, we analyzed EEC subtypes in the CHGA$^+$ cell population from different mutant backgrounds (Fig. 4B and fig. S17, A and B). Additional knockout of *INSM1* or *SOX4* caused a mild reduction of EC cells (*TPH1*) in CHGA$^+$ cells while further suppressing L cells (*GCG* and *PYY*) and D cells (*SST*). Loss of *PAX4* directed a robust conversion of EC-biased EECs into all other EEC subtypes. Basal secretion of subtype hormones (L cell, GLP-1; M/X cell, Ghrelin), confirmed their functionality (Fig. 4C). We also performed a single knockout of *SOX4* and *PAX4* (fig. S3A, S16B and S18A). The SOX4$^{-/-}$ single knockout significantly repressed L cells (*GCG*), N cells (*NTS*), D cells (*SST*), and G cells (*GAST*), whereas MLN-expressing M/X cells were not affected, agreeing with findings in *Sox4*$^{-/-}$ mice (*12*) (fig. S18, B to D). The PAX4$^{-/-}$ single knockout repressed ECs while mildly up-regulating other EEC subtypes. Overall, *PAX4*, as a direct target of *ZNF800*, appeared responsible for the EC-biased EEC differentiation trajectory seen upon loss of *ZNF800*.

Mouse ECs are exclusively *Pax4*-dependent (*45*), whereas *Arx* is essential for L, G, I, and N cells (*10*, *45*). We therefore probed a possible interaction between PAX4 and ARX in our organoids. ARX expression was down-regulated in ZNF800$^{-/-}$ organoids and up-regulated in

PAX4$^{-/-}$ single-knockout and ZNF800$^{-/-}$;PAX4$^{-/-}$ double-knockout organoids (Fig. 4D and fig. S19, A and B). Reintroduction of PAX4 in ZNF800$^{-/-}$;PAX4$^{-/-}$ organoids significantly repressed ARX expression (fig. S20, A and B). Furthermore, as expected, the EEC differentiation trajectory in ZNF800$^{-/-}$;PAX4$^{-/-}$ organoids was redirected to an EC-biased phenotype upon PAX4 reexpression.

Reciprocal transcriptional repression exists between *Pax4* and *Arx* during mouse pancreatic development (*46*, *47*). We assessed whether ARX also inhibits intestinal PAX4 expression by overexpressing ARX in ZNF800$^{-/-}$ organoids (fig. S21A). ARX overexpression was insufficient to reverse the EC-biased EEC cell type specification induced by the ZNF800-PAX4 axis (fig. S21, B and C). Furthermore, PAX4 expression was not affected upon ARX overexpression in the ZNF800$^{-/-}$ condition (fig. S21D). We next performed ChIP-qPCR of ARX on the PAX4 locus using FLAG-tagged ARX in human SI organoids (fig. S21E). Unlike ZNF800, ARX did not bind to the PAX4 upstream enhancer region (fig. S21F). We also performed ChIP-qPCR of PAX4 on the ARX locus using FLAG-tagged PAX4 (fig. S22A), focusing on six ultraconserved enhancers (*48*–*50*) (fig. S22B). PAX4 did not bind across these enhancers, including hs121, which exhibits PAX4 binding activity in the mouse pancreas (fig. S22C) (*49*). Overall, our results strongly supported that PAX4 unilaterally inhibits ARX, whereas the inhibitory effect does not involve direct chromatin interaction through the previously described regulatory elements.

Discussion

The human SI epithelium consists of at least 14 main cell types, including six EEC lineages (*51*, *52*). Leveraging the near-physiological cell heterogeneity of human SI organoids, we performed a CRISPR screen for positive and negative TF regulators of EEC lineage commitment. Our findings define a ZNF800-repressive mechanism upstream of the classic endocrine gene regulatory network (Fig. 4E). Among its direct downstream targets, *INSM1*, *SOX4*, and *NEUROG3* are well described central players that drive early EEC commitment. We also found that PAX4, in the absence of ZNF800, drives an EC-specific cell fate by suppressing differentiation of all other EEC subtypes.

The *PAX4* knockout rescued somatostatin-producing D cells in ZNF800$^{-/-}$ organoids (Fig. 2D and fig. S7A). Although *Pax4* controls both β- and somatostatin-producing δ-cell specification in the mouse pancreas (*46*), our single *PAX4* knockout had no effect on SST-expressing D cells (fig. S18C). This prompts questions into the generalizability of endocrine fate regulation between different digestive organs. Double depletion of *Pax4* and *Arx* promoted δ-cell specification in mice (*49*), suggesting that a third TF is involved in δ-cell cell differentiation, and that *Pax4* is likely to induce a β-cell fate at the expense of δ cells. Therefore, the up-regulation of *PAX4* by ZNF800$^{-/-}$ in human SI organoids might drive a further binary cell–fate decision between ECs and D cells.

ZNF800 is ubiquitously expressed along the crypt-villus axis in the adult human gut epithelium (fig. S4A), marked by H3K4me3 enrichment and low H3K27me3 levels (fig. S4B), indicating an open chromatin state and active expression. The scRNA-seq dataset of the developing human gut revealed that ZNF800 is already expressed at the earliest assayed time point of 6.1 postconception weeks (*53*). Given the downstream TF network described in our study, it appears likely that ZNF800 also plays an important function in embryonic development. A recent coexpression network study in mouse pancreatic development (*54*) revealed that expression of the mouse ortholog *zfp800* correlates with endocrine specification from embryonic day 8 (E8) until E15.5. Global knockout of *zfp800* caused postnatal lethality and disrupted early pancreatic development (including both endocrine and acinar cells at E18.5). Thus, the constitutive null phenotype hindered a mechanistic study of mouse *zfp800* function in the endocrine lineages. It is conceivable that ZNF800 might regulate β-cell differentiation in the human pancreas through the TF network described in this study.

REFERENCES AND NOTES

1. N. Barker *et al*., *Nature* **449**, 1003–1007 (2007).
2. Y. Sei, X. Lu, A. Liou, X. Zhao, S. A. Wank, *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G345–G356 (2011).
3. S. J. A. Buczacki *et al*., *Nature* **495**, 65–69 (2013).
4. O. Basak *et al*., *EMBO J.* **33**, 2057–2068 (2014).
5. O. Basak *et al*., *Cell Stem Cell* **20**, 177–190.e4 (2017).
6. H. Mutoh *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3560–3564 (1997).
7. F. J. Naya *et al*., *Genes Dev.* **11**, 2323–2334 (1997).
8. M. S. Gierl, N. Karoulias, H. Wende, M. Strehle, C. Birchmeier, *Genes Dev.* **20**, 2465–2478 (2006).
9. S. Desai *et al*., *Dev. Biol.* **313**, 58–66 (2008).
10. A. Du *et al*., *Dev. Biol.* **365**, 175–188 (2012).
11. S. Gross *et al*., *Development* dev.130682 (2016).
12. H. Gehart *et al*., *Cell* **176**, 1158–1173.e16 (2019).
13. T. Sato *et al*., *Nature* **459**, 262–265 (2009).
14. T. Sato *et al*., *Gastroenterology* **141**, 1762–1772 (2011).
15. G. Gradwohl, A. Dierich, M. LeMeur, F. Guillemot, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1607–1611 (2000).
16. M. Jenny *et al*., *EMBO J.* **21**, 6338–6347 (2002).
17. J. Lee *et al*., *eLife* **2**, e00940 (2013).
18. J. Beumer *et al*., *Cell* **181**, 1291–1306.e19 (2020).
19. G.-W. He *et al*., *Cell Stem Cell* **29**, 1333–1345.e6 (2022).
20. J. G. Doench *et al*., *Nat. Biotechnol.* **34**, 184–191 (2016).
21. S. A. Lambert *et al*., *Cell* **172**, 650–665 (2018).
22. L. Lin *et al*., *Cell Rep.* **33**, 108426 (2020).
23. N. F. Shroyer, D. Wallis, K. J. T. Venken, H. J. Bellen, H. Y. Zoghbi, *Genes Dev.* **19**, 2412–2417 (2005).
24. T. K. Noah, B. Donahue, N. F. Shroyer, *Exp. Cell Res.* **317**, 2702–2710 (2011).
25. T. Nakaya *et al*., *Cancer Res.* **74**, 2882–2891 (2014).
26. V. Korinek *et al*., *Nat. Genet.* **19**, 379–383 (1998).
27. E. Batlle *et al*., *Cell* **111**, 251–263 (2002).
28. M. van de Wetering *et al*., *Cell* **111**, 241–250 (2002).
29. J. H. van Es *et al*., *Mol. Cell. Biol.* **32**, 1918–1927 (2012).
30. K. Minamide *et al*., *Sci. Rep.* **10**, 14639 (2020).
31. T. Sato *et al*., *Nat. Cell Biol.* **22**, 919–926 (2020).
32. A. D. Gracz *et al*., *Gastroenterology* **155**, 1508–1523.e10 (2018).
33. M. Bjerknes, H. Cheng, *Dev. Biol.* **345**, 49–63 (2010).
34. H. S. Najafabadi *et al*., *Nat. Biotechnol.* **33**, 555–562 (2015).
35. A. Sakaue-Sawano *et al*., *Cell* **132**, 487–498 (2008).
36. S. Aibar *et al*., *Nat. Methods* **14**, 1083–1086 (2017).
37. M. H. Sham *et al*., *Cell* **72**, 183–196 (1993).
38. J. Ghislain, C. Desmarquet-Trin-Dinh, P. Gilardi-Hebenstreit, P. Charnay, M. Frain, *Development* **130**, 941–953 (2003).
39. A. Desmazières, P. Charnay, P. Gilardi-Hebenstreit, *J. Biol. Chem.* **284**, 10831–10840 (2009).
40. S. L. Dunwoodie, D. Henrique, S. M. Harrison, R. S. P. Beddington, *Development* **124**, 3065–3076 (1997).
41. R. M. Henke, D. M. Meredith, M. D. Borromeo, T. K. Savage, J. E. Johnson, *Dev. Biol.* **328**, 529–540 (2009).
42. B. E. Bernstein *et al*., *Cell* **125**, 315–326 (2006).
43. P. Voigt, W.-W. Tee, D. Reinberg, *Genes Dev.* **27**, 1318–1338 (2013).
44. A. B. Osipovich *et al*., *Development* **141**, 2939–2949 (2014).
45. A. Beucher *et al*., *PLOS ONE* **7**, e36449 (2012).
46. P. Collombat *et al*., *Genes Dev.* **17**, 2591–2603 (2003).
47. P. Collombat *et al*., *Cell* **138**, 449–462 (2009).
48. G. Bejerano *et al*., *Science* **304**, 1321–1325 (2004).
49. P. Collombat *et al*., *Development* **132**, 2969–2980 (2005).
50. D. E. Dickel *et al*., *Cell* **172**, 491–499.e15 (2018).
51. R. Elmentaite *et al*., *Nature* **597**, 250–255 (2021).
52. J. Burclaff *et al*., *Cell. Mol. Gastroenterol. Hepatol.* **13**, 1554–1589 (2022).
53. R. Elmentaite *et al*., *Dev. Cell* **55**, 771–783.e5 (2020).
54. A. B. Osipovich *et al*., *Development* **148**, dev196964 (2021).
55. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
56. Y. Luo *et al*., *Nucleic Acids Res.* **48** (D1), D882–D889 (2020).

ACKNOWLEDGMENTS

We thank R. Sherwood (Harvard Medical School, Brigham and Women's Hospital, Division of Genetics) for providing the TFome sgRNA library in the CRISPR-v2-FE backbone. We thank D. Krueger, A. de Graaff, and the Hubrecht Imaging Centre for microscopy assistance and the Hubrecht Flow Cytometry Core Facility for flow cytometry analysis and cell sorting. We thank the Máxima Single Cell Genomics Facility and Utrecht Sequencing Facility for sequencing support. We thank the Microscopy CORE Lab of the Faculty of Health, Medicine, and Life Sciences of Maastricht University for its help in transmission electron microscopy imaging. We thank the ENCODE Consortium and the ENCODE production laboratories (B. Bernstein, Broad; J. Stamatoyannopoulos, UW; B. Ren, UCSD) for generating the related ChIP-seq public datasets. **Funding:** This work was supported by the Netherlands Organ-on-Chip Initiative, a Dutch Research Council (NWO) Gravitation project (024.003.001) funded by the Ministry of Education, Culture, and Science of the government of the Netherlands (H.C.); the Oncode Institute (partly financed by the Dutch Cancer Society) (H.C. and J.H.v.E.); and the European Research Council under ERC advanced grant no. 101020405 (GutHormones) (H.C.). The project Organoids in time with project no. 2019.085 of the research program NWO Investment Large is financed by the NWO (H.C. S.v.d.B.). **Author contributions:** Conceptualization: L.L. and H.C.; Methodology: L.L., J.De., D.W., G.J.F.v.S., R.v.d.L., H.B., J.K., S.v.d.B., A.A.-R., C.L.-I., W.J.v.d.W., A.B., T.M., M.v.d.W., P.J.P., J.Dr., J.H.v.E., and H.C.; Investigation: L.L., J.De., G.J.F.v.S., H.B., J.K., A.A.-R., C.L.-I., and W.J.v.d.W.; Visualization: L.L., J.De., G.J.F.v.S., and C.L.-I.; Funding acquisition: J.H.v.E. and H.C.; Project administration: J.H.v.E. and H.C.; Supervision: J.H.v.E. and H.C.; Writing – original draft: L.L. and H.C.; Writing – review and editing: all authors. **Competing interests:** H.C. is an inventor on patents held by the Royal Netherlands Academy of Arts and Sciences that cover organoid technology. He is currently head of pharma Research and Early Development (pRED) at

Roche, Basel, Switzerland. H.C.'s full disclosure is given at https://www.uu.nl/staff/JCClevers/. **Data and materials availability:** Further information and requests for resources and reagents should be directed to the corresponding author. Specific and stable reagents generated in this study are available and can be requested from the corresponding author; a completed Materials Transfer Agreement may be required. Sharing organoid lines used in this study requires approval by our local institutional review board. CRISPR screen sequencing data, scRNA-seq data, and ChIP-seq data from this study have been deposited to the Gene Expression Omnibus (GEO) under accession no. GSE229586.

ChIP-seq data for histone marks H3K4me3 and H3K27me3 from human small intestine and colon were obtained from the ENCODE portal. We downloaded the ChIP-seq datasets from the ENCODE portal (*55*, *56*) (https://www.encodeproject.org/) with the following identifiers: ENCFF661MCT, ENCFF988AAQ, ENCFF219ZIO, ENCFF284XHG, ENCFF603QDZ, ENCFF825TAZ, ENCFF745IGA, and EENCFF830NSP. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi2246
Materials and Methods
Figs. S1 to S22
References (*57*–*76*)
Tables S1 to S8
MDAR Reproducibility Checklist

Submitted 17 April 2023; resubmitted 9 August 2023
Accepted 8 September 2023
10.1126/science.adi2246

ORGANIC CHEMISTRY

Multiplicative enhancement of stereoenrichment by a single catalyst for deracemization of alcohols

Lu Wen†, Jia Ding†, Lingfei Duan†, Shun Wang, Qing An, Hexiang Wang, Zhiwei Zuo*

Stereochemical enrichment of a racemic mixture by deracemization must overcome unfavorable entropic effects as well as the principle of microscopic reversibility; recently, photochemical reaction pathways unveiled by the energetic input of light have led to innovations toward this end, most often by ablation of a stereogenic C(sp^3)–H bond. We report a photochemically driven deracemization protocol in which a single chiral catalyst effects two mechanistically different steps, C–C bond cleavage and C–C bond formation, to achieve multiplicative enhancement of stereoinduction, which leads to high levels of stereoselectivity. Ligand-to-metal charge transfer excitation of a titanium catalyst coordinated by a chiral phosphoric acid or bisoxazoline efficiently enriches racemic alcohols that feature adjacent and fully substituted stereogenic centers to enantiomeric ratios up to 99:1. Mechanistic investigations support a pathway of sequential radical-mediated bond scission and bond formation through a common prochiral intermediate and reveal that, although the overall stereoenrichment is high, the selectivity in each individual step is moderate.

Ubiquitous C(sp^3)–C(sp^3) bonds in organic compounds have recently been exploited as unconventional functional handles for rapid complexity generation through skeletal editing (*1*–*5*). This logic presents intriguing opportunities for stereogenic bond construction via asymmetric catalysis. Most commonly, stereogenic C–C bonds are formed by facially selective addition to a prochiral starting material (Fig. 1A) (*6*–*8*). Critically, the overall enantioselectivity of this process is thus a direct function of the stereodifferentiation of the two prochiral faces in the single stereodefining step [enantiomeric ratio (er) = k_R/k_S, where k_R is the rate of formation of the (*R*)-enantiomer and k_S is the rate of formation of the (*S*)-enantiomer]. As a result, exceptionally high degrees of stereoinduction are required in this irreversible addition step to achieve synthetically useful results ($k_R >> k_S$). An alternative paradigm can be envisioned transiting through a prochiral intermediate to enable a formally reversible C–C bond formation process, which can convert a mixture of racemates into enantiomerically pure compounds (Fig. 1B). This approach would enable a broader purview for catalytic deracemization because C–C bonds constitute the fundamental three-dimensional skeleton of complex molecules. Through stereoisomerization of the molecular core, as opposed to the peripheral C–H bonds, consecutive or fully substituted stereogenic centers could be directly enriched by using this cycle of stereocenter-ablating C–C bond scission and stereocenter-generating C–C bond formation, which would provide a powerful platform for enantioenrichment via catalytic deracemization.

Within this framework, we recently wondered whether one catalyst could be directly responsible for each of these stereoselective steps by means of different mechanisms, which would thus allow for deracemization. The multiplicative nature of stereoinduction would lead to amplification of the enantiodifferentation ability of the catalyst, in analogy to the Horeau principle in which multiple asymmetric steps do not necessarily need rigid stereocontrol to obtain highly enantioenriched diastereomers (*9*–*11*). However, in this case, because one of the stereodifferentiating steps in the catalytic cycle involves stereoablation, the same multiplicative enhancement could be achieved in the context of enantioinduction of even a single stereocenter (er = $k_R k_{-S}/k_S k_{-R}$). To achieve synthetically useful levels of enantioinduction [>90% enantiomeric excess (ee)], the single catalyst would only be required to achieve roughly 60% ee in each of the stereocenter-ablating and stereocenter-generating steps.

Although conceptually simple and synthetically intriguing, the development of such processes has been hindered by the principle of microscopic reversibility (*12*–*18*). To address this critical issue, previous studies have opted to use a single chiral catalyst for deracemization through a two-step process that consists of a nonselective step and a stereodifferentiating step. Notable early studies on redox-driven deracemizations by Toste *et al.* (*19*) and Zhou *et al.* (*20*), as well as recent breakthroughs in photochemical deracemization such as Bach *et al.*'s seminal works that used enantioselective triplet sensitization (*18*, *21*–*23*) and several other elegant systems with reversible C–H bond formation (*24*–*29*), have demonstrated the effectiveness of this approach (*30*, *31*). Very recently, Knowles and Miller and co-workers used an elegantly designed tricatalytic system that consisted of one chiral base for C–H bond breakage and a chiral thiol catalyst for C–H bond formation, which exploited the synergistic effect of two chiral catalysts for the photocatalytic deracemization of cyclic ureas (*32*). It is important to note that using one chiral catalyst to effect asymmetric induction in both bond-breaking and bond-making events will result in an overall equilibrium of racemate because of the shared energy surface (*33*, *34*), unless distinct reaction pathways can be realized by a single catalyst.

Recently, we have applied ligand-to-metal charge transfer (LMCT) excitation in an orchestrated sequence of in situ coordination, LMCT homolysis, and alkoxy radical–mediated β-scission to the catalytic activation of C–C bonds of free alcohols for fragment couplings (*35*–*38*). Driven by the thermodynamic stability of radicals, both strained and unstrained C–C bonds can be selectively and irreversibly cleaved into a carbonyl unit and a carbon-centered radical (*39*). Notably, high-valent metal ions, including commonly used Lewis acids, can be directly used for LMCT catalysis (*40*–*45*), which led us to consider the possibility of photocatalytic deracemization using chiral ligand–coordinated, LMCT-competent Lewis acid catalysts (Fig. 1C). Conceivably, asymmetric LMCT catalysts could leverage the chiral recognition of racemic alcohols in the coordination step to preferentially form one diastereomeric metal alkoxide complex, initiating enantioselective C–C bond scission, even if the β-scission process may not respond to asymmetric induction. The concurrent generation of a carbonyl fragment and a transient carbon-centered radical would set the stage for single-electron reduction and enantioface-differentiating addition to rebuild the stereogenic C–C bond, which could be facilitated by the lower-valent chiral metal complex generated in the photoexcitation event (*46*–*55*). In this work, we realize this design plan and describe an LMCT-enabled deracemization platform that uses a single chiral Ti-catalyst to induce decoupled and enantioselective C–C bond cleavage and formation, which results in high stereoselectivity through multiplicative enhancement of stereoinduction. By exploiting the photocatalytic properties of a common Lewis acidic Ti(IV) catalyst ligated by chiral phosphoric acid or bisoxazoline, racemic alcohols—including those that feature adjacent and fully substituted stereogenic centers—can be efficiently converted into their enantioenriched forms with pronounced selectivity.

Catalyst optimization with cyclic alcohols

We selected the cis isomer of 2-phenylcyclopentanol (**1**) as the model substrate to explore LMCT catalysis for deracemization (Fig. 2). Extensive evaluation of high-valent metal catalysts

State Key Laboratory of Organometallic Chemistry, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200032, China.
*Corresponding author. Email: zuozhw@sioc.ac.cn
†These authors contributed equally to this work.

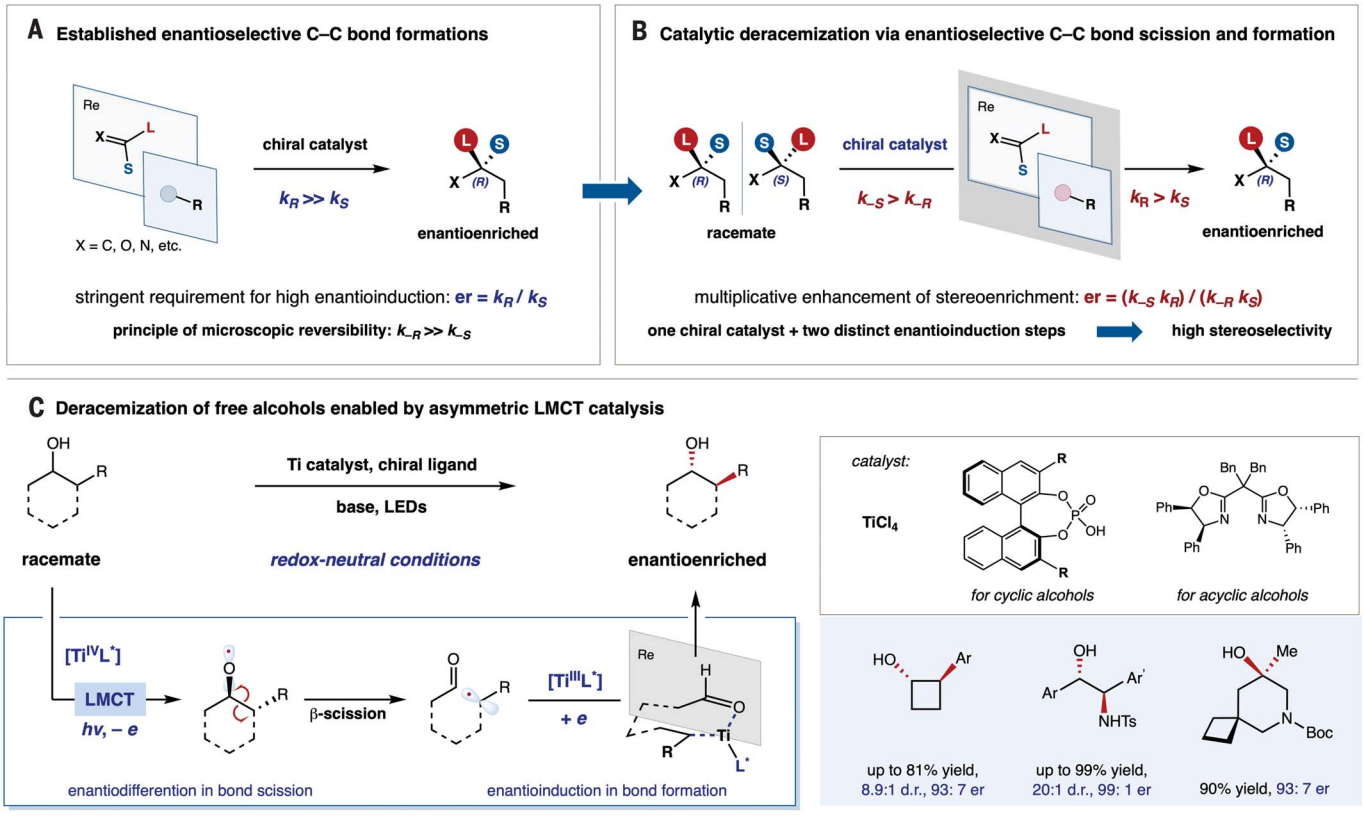
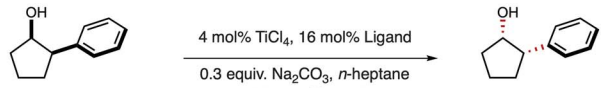
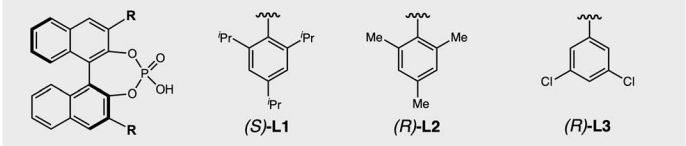


Fig. 1. Catalytic deracemization paradigms via enantioselective bond scission and formation. Multiplicative enhancement of two distinct enantioinduction steps enables high levels of stereoselectivities in deracemization. (**A**) Established enantioselective C–C bond formations. (**B**) Catalytic deracemization via enantioselective C–C bond scission and formation. (**C**) Deracemization of free alcohols enabled by asymmetric LMCT catalysis. Bn, benzyl; Boc, *t*-butyloxy carbonyl; *hv*, photon energy; L*, chiral ligand; Me, methyl; Ts, *p*-toluenesulfonyl; Ph, phenyl; Re, *R* prochiral face.



entry	ligand	yield (%)	d.r.	er
1	(*S*)-**L1**	93	8.8:1	99:1
2	(*R*)-**L2**	94	2.2:1	13:87
3	(*R*)-**L3**	96	1:1	62:38
4	(*R*)-**L4**	99	1:1	41:59
5	(*S*)-**L5**	97	1.6:1	44:56
6	(*R*)-**L6**	94	1.2:1	66:34
7	(*R*)-**L7**	99	3.2:1	21:79
8	(*S*)-**L1**, $EtN(^iPr)_2$ instead of Na_2CO_3	99	3.9:1	93:7
9	(*R*)-**L1**	99	8.8:1	1:99



entry	control experiments of entry 1	yield (%)	d.r.	er
10	$CeCl_3$ instead of $TiCl_4$	97	>20:1	50:50
11	$FeCl_3$ instead of $TiCl_4$	95	>20:1	50:50
12	no (*S*)-**L1**	95	1:1.2	50:50
13	no $TiCl_4$	>99	>20:1	50:50
14	dark	>99	>20:1	50:50

Fig. 2. Reaction optimization. Reactions were performed on a 0.05-mmol scale. Yields, dr, and er (given for major diastereomers) were determined by high-performance liquid chromatography (HPLC) analysis. C_6F_5, pentafluorophenyl; $EtN(^iPr)_2$, diisopropylethylamine; Tf, trifluoromethanesulfonyl; TRIP, 2,4,6-triisopropylphenyl.

with chiral ligands under the irradiation of light-emitting diodes (LEDs) (peak wavelength, 395 nm; light intensity, 0.8 W/cm^2) at 20°C (see fig. S1 for detailed experimental setup) revealed the combination of $TiCl_4$ and chiral phosphoric acid (CPA) as an effective catalyst combination for photocatalytic deracemization. This chiral Lewis acid combination has been previously used by Leibfarth to control enantiodifferentiating C–C bond formation for stereoselective cationic polymerization (*56*). In practice, we found that the combination of 4 mol % $TiCl_4$ and 16 mol % (*S*)-**L1** in the presence of substoichiometric Na_2CO_3 achieved optimal efficiency and selectivity, generating

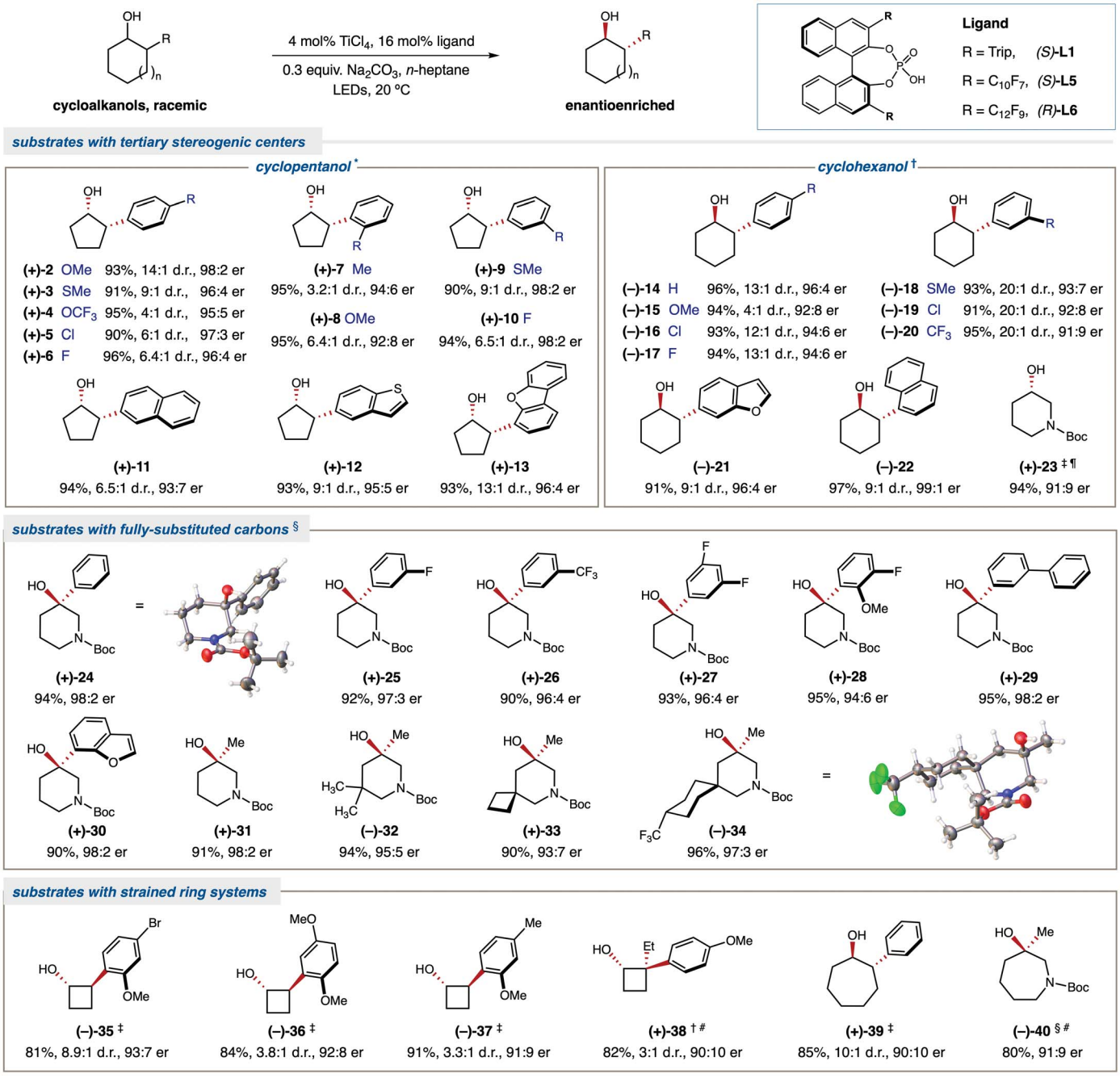

cycloalkanols, racemic
4 mol% TiCl4, 16 mol% ligand
0.3 equiv. Na2CO3, n-heptane
LEDs, 20 °C
enantioenriched
Ligand
R = Trip, (S)-L1
R = C10F7, (S)-L5
R = C12F9, (R)-L6
substrates with tertiary stereogenic centers
cyclopentanol*
cyclohexanol†
(+)-2 OMe 93%, 14:1 d.r., 98:2 er
(+)-3 SMe 91%, 9:1 d.r., 96:4 er
(+)-4 OCF3 95%, 4:1 d.r., 95:5 er
(+)-5 Cl 90%, 6:1 d.r., 97:3 er
(+)-6 F 96%, 6.4:1 d.r., 96:4 er
(+)-7 Me 95%, 3.2:1 d.r., 94:6 er
(+)-8 OMe 95%, 6.4:1 d.r., 92:8 er
(+)-9 SMe 90%, 9:1 d.r., 98:2 er
(+)-10 F 94%, 6.5:1 d.r., 98:2 er
(+)-11 94%, 6.5:1 d.r., 93:7 er
(+)-12 93%, 9:1 d.r., 95:5 er
(+)-13 93%, 13:1 d.r., 96:4 er
(–)-14 H 96%, 13:1 d.r., 96:4 er
(–)-15 OMe 94%, 4:1 d.r., 92:8 er
(–)-16 Cl 93%, 12:1 d.r., 94:6 er
(–)-17 F 94%, 13:1 d.r., 94:6 er
(–)-18 SMe 93%, 20:1 d.r., 93:7 er
(–)-19 Cl 91%, 20:1 d.r., 92:8 er
(–)-20 CF3 95%, 20:1 d.r., 91:9 er
(–)-21 91%, 9:1 d.r., 96:4 er
(–)-22 97%, 9:1 d.r., 99:1 er
(+)-23‡¶ 94%, 91:9 er
substrates with fully-substituted carbons§
(+)-24 94%, 98:2 er
(+)-25 92%, 97:3 er
(+)-26 90%, 96:4 er
(+)-27 93%, 96:4 er
(+)-28 95%, 94:6 er
(+)-29 95%, 98:2 er
(+)-30 90%, 98:2 er
(+)-31 91%, 98:2 er
(–)-32 94%, 95:5 er
(+)-33 90%, 93:7 er
(–)-34 96%, 97:3 er
substrates with strained ring systems
(–)-35‡ 81%, 8.9:1 d.r., 93:7 er
(–)-36‡ 84%, 3.8:1 d.r., 92:8 er
(–)-37‡ 91%, 3.3:1 d.r., 91:9 er
(+)-38†# 82%, 3:1 d.r., 90:10 er
(+)-39‡ 85%, 10:1 d.r., 90:10 er
(–)-40§# 80%, 91:9 er

Fig. 3. Reaction scope of cycloalkanols. Yields and dr and er values were determined based on isolated products and were the average of three parallel reactions. *4 mol % $TiCl_4$ with 16 mol % **L1**. †2 mol % $TiCl_4$ with 8 mol % **L6**. ‡4 mol % $TiCl_4$ with 16 mol % **L6**. §2 mol % $TiCl_4$ with 8 mol % **L5**. ¶–40°C. #0°C. See supplementary materials for experimental details. $C_{10}F_7$, heptafluoronaphthyl; $C_{12}F_9$, nonafluorobiphenyl; Et, ethyl.

cis (1*S*,2*S*)-**1** with high efficiency and notable enantioselectivity (93% yield, 99:1 er, entry 1). Under these mild and redox-neutral conditions, only a small amount of trans isomer was obtained along with the desired cis isomer [8.8:1 diastereomeric ratio (dr)], without the observation of side products such as aldehyde from photocatalytic ring-opening processes, ketone from 2e oxidations, or alkene from dehydration (see fig. S2). The steric environment of the CPA has a notable influence on the stereochemical outcome. Replacing the sterically hindered tri-*iso*-propylphenyl rings of **L1** with trimethylphenyl (**L2**) resulted in markedly diminished stereoselectivity (entry 2), whereas phosphoric acids bearing electron-deficient arene substituents (**L3**, **L4**, **L5**, or **L6**) resulted in rather low selectivities (entries 3 to 6). Use of the more acidic but less coordinating *N*-triflyl phosphoramide **L7** in place of **L1** resulted in markedly lower stereoselectivity (3.2:1 dr, 21:79 er) despite similar steric environments (entry 7). Organic base diisopropylethylamine was also found to be effective for the desired deracemization (entry 8). Changing the loading of CPA relative to Ti from 4:1 to 2:1 had little effect on the efficiency and stereoselectivity (see table S3), and the 4:1 ratio was adopted for best reproducibility on small-scale parallel reactions. A linear correlation between the enantioenrichment observed in the reaction and the ee value of CPA was identified (see fig. S4), and the absence

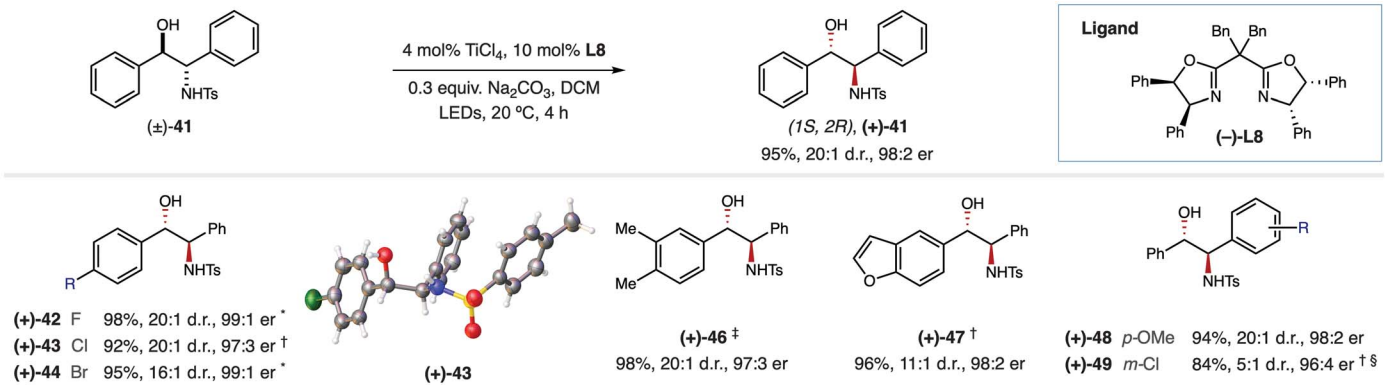


Fig. 4. Deracemization of acyclic alcohols. Yields and dr and er values were determined on the basis of isolated products and were the average of three parallel reactions. *8 mol % $TiCl_4$ with 20 mol % **L8**. †12 mol % $TiCl_4$ with 30 mol % **L8**. ‡Dichloromethane (DCM) and benzotrifluoride ($PhCF_3$) (1:1). §$PhCF_3$ instead of DCM.

of a nonlinear effect in this deracemization indicated that a single CPA is presumably involved at the Ti center (*57*). Use of (*R*)-**L1** resulted in identical values of stereoselectivity with opposite configuration, supporting the catalyst-controlled stereoselectivity (entry 9). Furthermore, deracemization of the secondary alcohol was not observed when swapping $TiCl_4$ with $CeCl_3$ or $FeCl_3$ (entries 10 and 11), and instead, a trace amount of 5-phenylpentanal was observed, indicating that Ce and Fe LMCT catalysts are capable of C–C bond scission but not effective at promoting C–C bond–forming addition. Moreover, control experiments revealed that **L1**, $TiCl_4$, and an LED light are essential for the deracemization (entries 12 to 14).

As demonstrated in Fig. 3, this Ti-LMCT–enabled photocatalytic deracemization can be carried out on structurally diverse cycloalkanols of varying ring sizes, indicating that ring strain is not a decisive factor. In cyclopentanol and cyclohexanol substrates, we found that the deracemization is somewhat insensitive to the electronic nature of the β-aromatic rings. Excellent enantioselectivities could be obtained for a series of diverse substituents on the para, ortho, and meta positions, with trivial differences in the diastereoselectivities. Even in the presence of large aromatic rings such as naphthalene (**11**), benzothiophene (**12**), and dibenzofuran (**13**), the thermodynamically less-stable cis isomers were preferentially formed for cyclopentanols. Regarding the six-membered ring system, enantioenriched trans isomers were selectively generated, presumably by virtue of a chair-like cyclic transition state for C–C bond formation. We found that (*R*)-**L6** was most effective for the deracemization of trans-2-substituted cyclohexanols. Notably, enantioenriched 3-hydroxy piperidine (**23**), an essential building block for the syntheses of important marketed pharmaceuticals such as zamifenacin, piperidolate, and benidipine, can also be selectively obtained with this straightforward protocol (*58*). Critically, enantioenriched synthesis of α-tertiary alcohols with fully substituted stereogenic carbons could be facilely achieved by this deracemization platform. With the catalytic combination of $TiCl_4$ and (*S*)-**L5**, a variety of tertiary 3-hydroxy piperidines could be obtained with high levels of enantioenrichment, regardless of the electronic nature of the α-aromatic rings. Tertiary alcohols with α-methyl substituents can also be accommodated, which renders structurally diversified piperidine scaffolds with excellent stereoselectivity. Further studies revealed that larger alkyl groups at the α-position, including ethyl and isobutyl groups, resulted in somewhat lower er (see table S10). The stereochemistry of tertiary alcohols **24** and **34** has been unambiguously assigned by single-crystal x-ray diffraction. Even β-substituted cyclobutanols were compatible with this deracemization paradigm and generated enantioenriched cyclobutanols **35** to **38** with good enantioselectivities through enantioselective C–C bond scission and reformation of the highly strained four-membered ring. The generation of aldehyde byproducts through a photocatalytic ring-opening process was observed in those strained substrates, which resulted in declined yields of deracemization. Secondary and tertiary alcohols embedded in seven-membered carbocycle or azepane scaffolds were also compatible, albeit with moderate selectivities. We conducted a scaled-up deracemization of cyclohexanol **14** at 1 mmol scale by using a simple and easily assembled photoflow system (see section 5.1 in the supplementary materials), which rendered (–)-**14** with similar levels of selectivities as the standard conditions (91% yield, 7.2:1 dr, 95:5 er).

Extension to acyclic alcohols

This protocol is not limited to cyclic hydroxylated scaffolds because the deracemization of acyclic alcohols has been validated on 1,2-diaryl aminoalcohols, a privileged scaffold for chiral ligand synthesis (Fig. 4). Under similar catalytic conditions as for the cyclic substrates, with chiral bisoxazoline **L8** as the ligand (see table S11 for the ligand evaluation), enantioenriched (+)-**41** can be obtained with excellent yield and stereoselectivity (20:1 dr, 98:2 er). A variety of substitution patterns on the α- and β-aromatic rings proved compatible, and high yields and excellent stereoselectivities have been obtained with no sensitivity to the electronic property of the aromatic rings. This deracemization has provided a practical approach to enantioenriched aminoalcohols with electronically differentiated aryl rings on two different carbon termini (*59*, *60*). Moreover, we conducted 1-mmol-scale reactions with **41** and **43**, which delivered enantiopure alcohols with identical selectivities to the standard condition at prolonged reaction time (10 hours). Compound (+)-**43** can be elaborated via deprotection and double condensation to generate a differentially arylated bisoxazoline ligand, the para-chloro substituted version of **L8** (see section 5.2 in the supplementary materials).

Mechanistic investigations

We carried out mechanistic investigations to elucidate the reaction pathway. Using C–H-deuterated cylopentanol **50** and aminoalcohol **51** delivered nearly identical stereochemical outcomes as the unlabeled analogs without eroding the deuteration ratio, which precludes hydrogen atom transfer (HAT) or stepwise oxidation reduction as possible deracemization pathways (Fig. 5A). With the same ligand, enantioconvergent transformations that used the trans isomer of **1** or the opposite enantiomer (1*R*,2*R*)-**1** resulted in the same (1*S*,2*S*)-**1** product with identical stereoselectivity, supporting a common, prochiral intermediate generated by the C–C bond scission (Fig. 5B). This conclusion can also be drawn from the same set of experiments with aminoalcohol **41** (fig. S8).

The LMCT-homolysis behavior of Ti(IV) complexes has been validated by steady-state photolysis experiments by using an in situ–generated putative complex [Ti(IV)L(OR)] (Fig. 5C). We used a mononuclear [Ti^{IV}(L8)Cl_4] complex, characterized by x-ray diffraction, in the ligand

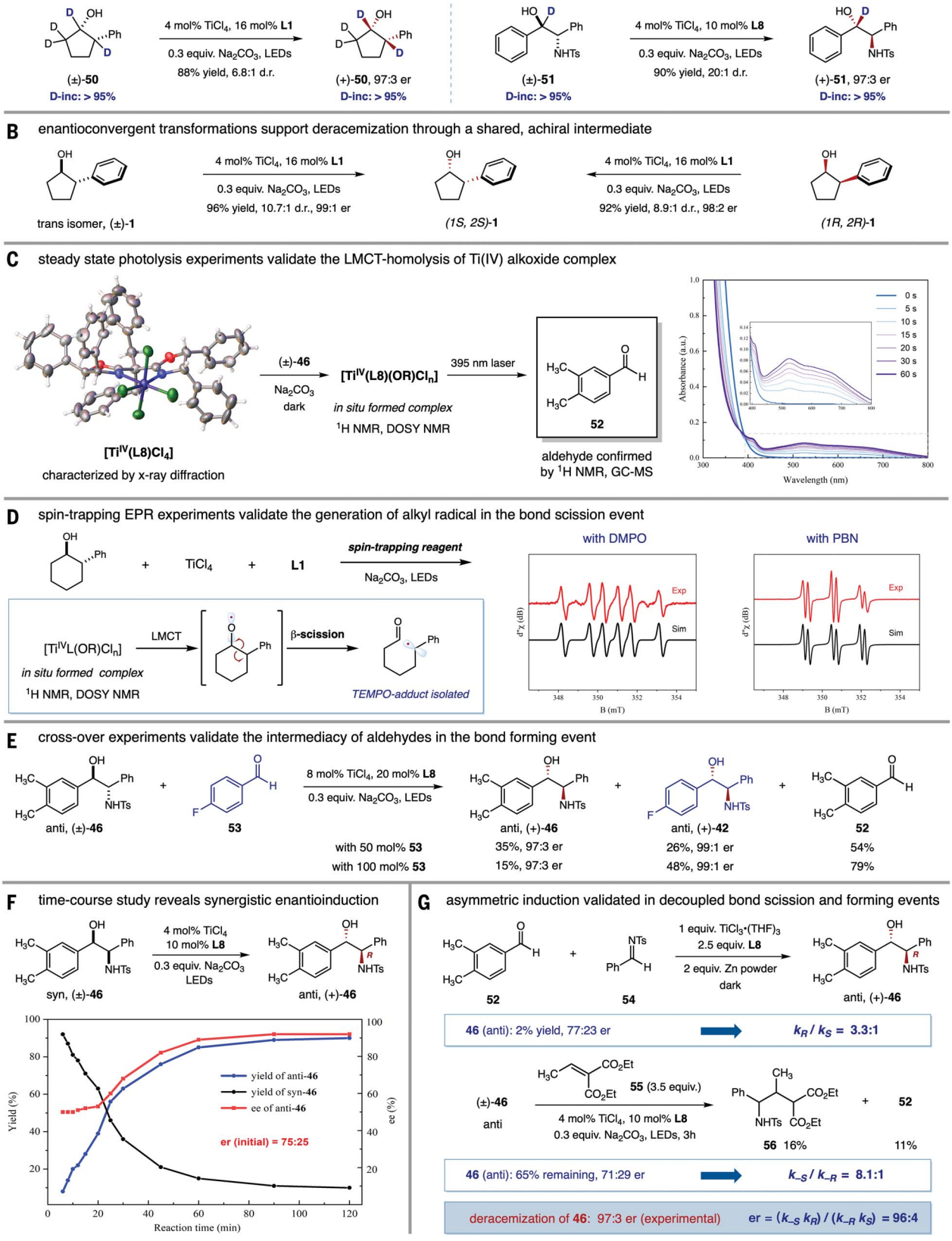


Fig. 5. Preliminary mechanistic investigations. (**A**) Isotope labeling experiments. (**B**) Enantioconvergent transformations. (**C**) Steady-state photolysis experiments. a.u., arbitrary unit; GC-MS, gas chromatography–mass spectrometry. (**D**) Spin-trapping EPR experiments. Exp, experimental; Sim, simulation. (**E**) Crossover experiment. (**F**) Time-course study. (**G**) Decoupled radial-mediated bond formation and photocatalytic bond scission. THF, tetrahydrofluoran.

exchange with an equimolar amount of **46** for the in situ generation of [Ti^{IV}(L8)(OR)]. The ^{1}H nuclear magnetic resonance (NMR) spectrum displayed characteristic signals that can be assigned to the bisoxazoline **L8** and alkoxide ligands in a 1:1 ratio, and diffusion ordered spectroscopy (DOSY) NMR experiments confirmed that the obtained species are monomeric complex and not higher aggregates (fig. S10). In the same vein, the putative complex generated from $TiCl_4$ with an equimolar amount of **L1** and **14** was prepared (fig. S12). Both complexes exhibited substantial absorptions in the proximity of 395 nm. Under the irradiation of a 395-nm laser, with the gradual decay of the absorption band ranging from 300 to 400 nm, a weakly absorbing band in the 500- to 800-nm region started to emerge as the originally light yellow solution turned blue, which is indicative of the generation of Ti(III) species (*43*). Upon exposure to air, the blue solution was swiftly bleached to the yellow Ti(IV) complex with the ultraviolet to visible spectrum fully recovered. Aldehyde **52**, formed by radical-mediated bond scission, was detected by gas chromatography–mass spectrometry and ^{1}H NMR, supporting the generation of alkoxy radicals. Regarding the cycloalkanols, the proposed alkyl radical generation by rapid alkoxy radical β-scission was monitored by operando electron paramagnetic resonance experiments with 5,5-dimethyl-1-pyrroline *N*-oxide (DMPO) and α-phenyl-*N*-*tert*-butylnitrone (PBN) as spin-trapping reagents. As depicted in Fig. 5D, the characteristic signals of alkyl radical adduct with DMPO [hyperfine splitting constant A_N = 14.4 G, A_H = 20.9 G, spectroscopic splitting factor (g) = 2.0061] or PBN (A_N = 14.6 G, A_H = 2.4 G, g = 2.0062) can be clearly distinguished, whereas alkoxy radical adduct was not detected, indicating a fast β-scission process that favors the C–C bond cleavage or a concerted oxidation with C–C bond scission (*61*, *62*). Moreover, we further confirmed the alkyl radical generated in the β-scission by the isolation and characterization of the 2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPO) adduct.

We rationalized that within the same mechanistic framework, the aromatic aldehyde generated from the β-scission process of acyclic aminoalcohols could be exchanged with a more reactive exogenous aldehyde for the C–C bond formation event to generate another aminoalcohol product. Indeed, under the same catalytic conditions for deracemization, crossover experiments with racemic **46** and parafluorobenzaldehyde **53** resulted in a newly formed aminoalcohol **42** that contained a parafluorophenyl group (Fig. 5E). The newly generated aminoalcohol **42** underwent deracemization to reach the same level of enantioenrichment as the standard condition (99:1 er). The yield of **42** can be further enhanced at the expense of deracemization product **46** with increased concentration of aldehyde **53**. The foregoing results collectively and unambiguously support the intermediacy of aldehydes and carbon-centered radicals generated by β-scission as the shared intermediate in this deracemization paradigm. The high efficiency of the desired cyclobutane ring formation in the deracemization of cyclobutanols is notable because direct intramolecular radical addition to the carbonyl to form a four-membered ring is energetically unfavorable, considering the high exothermicity and low activation barrier of cyclobutoxy cleavage (*39*, *63*). There are two seemingly reasonable pathways for this cyclization that account for the ability of Ti(III) to mediate such a radical bond formation: (i) reductive capture of the formed alkyl radical by photogenerated Ti(III) preceding a stereodifferentiating 1,2-addition with the carbonyl group via a six-membered cyclic transition state; or (ii) reversible reduction of the carbonyl to a transient Ti-ketyl radical followed by radical-radical coupling. Although the relative 1,2-stereochemistry observed for cyclobutanol (trans), cyclopentanol (cis), and cyclohexanol (trans) substrates is reminiscent of pseudoequitorial positioning of bulky groups in cyclic transition states (fig. S18), the precise nature of this ring closure warrants further investigation.

Insight into the level of asymmetric induction incurred in the C–C bond–forming events can be gleaned by tracking the time course of the deracemization of the syn-**46**, which results in enantioenriched anti-**46** (Fig. 5F). In the first 20 min, the bond cleavage of syn-**46** and bond formation step to generate anti-**46** results in a steady er value of 75:25. As the concentration of newly generated anti-**46** approaches that of syn-**46**, the er value of anti-**46** gradually increases as the deracemization of the formed major diastereomer takes place, eventually reaching an equilibrium value of 97:3 er. The steady er value of anti-**46** during the initial 20 min indicates that the newly formed anti-**46** is not yet substantially involved in the bond cleavage cycle, and the enantiofacial-determining ratio of the bond formation step can be estimated as 75:25, which results in k_R/k_S ratio of 3:1. This approximate value is supported by a 77:23 er (k_R/k_S ratio of 3.3:1) obtained for the inefficient reductive coupling of aldehyde **52** and imine **54** under dark conditions with **L8**-ligated $TiCl_3$, despite predominant pinacol coupling (Fig. 5G) (*51*, *53*, *64*). Conversely, we could directly probe the bond scission of aminoalcohol **46** by the introduction of a large amount of radical trapping reagent **55**, which would lead to kinetic resolution based on enantioselective C–C bond cleavage. The asymmetric induction of the C–C bond scission step can be established as a k_{-S}/k_{-R} ratio of 8.1:1, with the (*S*)-enantiomer preferentially consumed. The predicated er value based on the multiplicative effect (er = $k_R k_{-S}/k_S k_{-R}$) of these two isolated processes is 96:4, in excellent agreement with the experimental value in the deracemization reaction (97:3) (see fig. S17 for a discussion on cyclic alcohol **1**).

The LMCT catalytic paradigm enables the use of one chiral Ti catalyst to induce two enantioselective events in a single catalytic cycle. The multiplicative enhancement of stereoinduction, validated in consecutive stereocenter-ablating C–C bond scission and stereocenter-generating C–C bond formation, enables high levels of stereoselectivities with two moderate enantioinductions. This mechanistic paradigm holds promise for the development of a wide array of asymmetric catalytic reactions through the statistical upgrading of stereoenrichment.

REFERENCES AND NOTES

1. G. Dong, N. Cramer, *C–C Bond Activation*, vol. 346 of *Topics in Current Chemistry*, G. Dong, Ed. (Springer, 2014).
2. L. Souillart, N. Cramer, *Chem. Rev.* **115**, 9410–9464 (2015).
3. B. Wang, M. A. Perea, R. Sarpong, *Angew. Chem. Int. Ed.* **59**, 18898–18919 (2020).
4. X.-Y. Yu, J.-R. Chen, W.-J. Xiao, *Chem. Rev.* **121**, 506–561 (2021).
5. J. Jurczyk *et al*., *Nat. Synth.* **1**, 352–364 (2022).
6. P. J. Walsh, M. C. Kozlowski, *Fundamentals of Asymmetric Catalysis* (University Science Books, 2009).
7. M. Shibasaki, M. Kanai, *Chem. Rev.* **108**, 2853–2873 (2008).
8. D. J. Ramón, M. Yus, *Chem. Rev.* **106**, 2126–2208 (2006).
9. J. P. Vigneron, M. Dhaenens, A. Horeau, *Tetrahedron* **29**, 1055–1059 (1973).
10. A. M. Harned, *Tetrahedron* **74**, 3797–3841 (2018).
11. M. Burns *et al*., *Nature* **513**, 183–188 (2014).
12. A. E. Wendlandt, *Science* **366**, 304–305 (2019).
13. Q. Shi, J. Ye, *Angew. Chem. Int. Ed.* **59**, 4998–5001 (2020).
14. M. Huang, T. Pan, X. Jiang, S. Luo, *J. Am. Chem. Soc.* **145**, 10917–10929 (2023).
15. J. Großkopf, T. Bach, *Angew. Chem. Int. Ed.*e202308241 (2023).
16. A. Tröster, A. Bauer, C. Jandl, T. Bach, *Angew. Chem. Int. Ed.* **58**, 3538–3541 (2019).
17. L. Wimberger, T. Kratz, T. Bach, *Synthesis* **51**, 4417–4416 (2019).
18. X. Li *et al*., *Angew. Chem. Int. Ed.* **59**, 21640–21647 (2020).
19. A. D. Lackner, A. V. Samant, F. D. Toste, *J. Am. Chem. Soc.* **135**, 14090–14093 (2013).
20. Y. Ji, L. Shi, M.-W. Chen, G.-S. Feng, Y.-G. Zhou, *J. Am. Chem. Soc.* **137**, 10496–10499 (2015).
21. A. Hölzl-Hobmeier *et al*., *Nature* **564**, 240–243 (2018).
22. M. Plaza, J. Großkopf, S. Breitenlechner, C. Bannwarth, T. Bach, *J. Am. Chem. Soc.* **143**, 11209–11217 (2021).
23. T. Kratz *et al*., *J. Am. Chem. Soc.* **144**, 10133–10138 (2022).
24. J. Großkopf *et al*., *J. Am. Chem. Soc.* **143**, 21241–21245 (2021).
25. R. J. Kutta *et al*., *J. Am. Chem. Soc.* **145**, 2354–2363 (2023).
26. C. Zhang *et al*., *J. Am. Chem. Soc.* **143**, 13393–13400 (2021).
27. M. Huang, L. Zhang, T. Pan, S. Luo, *Science* **375**, 869–874 (2022).
28. Z. Gu *et al*., *Angew. Chem. Int. Ed.* **61**, e202211241 (2022).
29. Q. Chen *et al*., *Chem. Sci.* **14**, 1715–1723 (2023).
30. M. Wan, S. Sun, Y. Li, L. Liu, *Angew. Chem. Int. Ed.* **56**, 5116–5120 (2017).
31. Z. Zhang, X. Hu, *Angew. Chem. Int. Ed.* **60**, 22833–22838 (2021).
32. N. Y. Shin, J. M. Ryss, X. Zhang, S. J. Miller, R. R. Knowles, *Science* **366**, 364–369 (2019).
33. D. G. Blackmond, *Angew. Chem. Int. Ed.* **48**, 2648–2654 (2009).
34. W. Kroutil, K. Faber, *Tetrahedron Asymmetry* **9**, 2901–2913 (1998).
35. J.-J. Guo *et al*., *Angew. Chem. Int. Ed.* **55**, 15319–15322 (2016).
36. A. Hu *et al*., *J. Am. Chem. Soc.* **140**, 13580–13585 (2018).
37. K. Zhang, L. Chang, Q. An, X. Wang, Z. Zuo, *J. Am. Chem. Soc.* **141**, 10556–10564 (2019).

38. Y. Chen, X. Wang, X. He, Q. An, Z. Zuo, *J. Am. Chem. Soc.* **143**, 4896–4902 (2021).
39. L. Chang, Q. An, L. Duan, K. Feng, Z. Zuo, *Chem. Rev.* **122**, 2429–2486 (2022).
40. S. M. Treacy, T. Rovis, *J. Am. Chem. Soc.* **143**, 2729–2735 (2021).
41. Z. Zhang *et al*., *Angew. Chem. Int. Ed.* **59**, 9355–9359 (2020).
42. M. Yamane, Y. Kanzaki, H. Mitsunuma, M. Kanai, *Org. Lett.* **24**, 1486–1490 (2022).
43. A. Fermi, A. Gualandi, G. Bergamini, P. G. Cozzi, *Eur. J. Org. Chem.* **2020**, 6955–6965 (2020).
44. F. Juliá, *ChemCatChem* **14**, e202200916 (2022).
45. C. Wegeberg, O. S. Wenger, *JACS Au* **1**, 1860–1876 (2021).
46. A. Gansäuer, L. Shi, M. Otte, *J. Am. Chem. Soc.* **132**, 11858–11859 (2010).
47. J. Streuff, M. Feurer, P. Bichovski, G. Frey, U. Gellrich, *Angew. Chem. Int. Ed.* **51**, 8661–8664 (2012).
48. W. Hao, J. H. Harenberg, X. Wu, S. N. MacMillan, S. Lin, *J. Am. Chem. Soc.* **140**, 3514–3517 (2018).
49. A. Gualandi *et al*., *ACS Catal.* **10**, 3857–3863 (2020).
50. F. Li *et al*., *Angew. Chem. Int. Ed.* **60**, 1561–1566 (2021).
51. F. Calogero *et al*., *Chem. Sci.* **13**, 5973–5981 (2022).
52. X. Peng *et al*., *J. Org. Chem.* **88**, 6333–6346 (2023).
53. B. Rossi, S. Prosperini, N. Pastori, A. Clerici, C. Punta, *Molecules* **17**, 14700–14732 (2012).
54. T. McCallum, X. Wu, S. Lin, *J. Org. Chem.* **84**, 14369–14380 (2019).
55. X. Wu, Y. Chang, S. Lin, *Chem* **8**, 1805–1821 (2022).
56. A. J. Teator, F. A. Leibfarth, *Science* **363**, 1439–1443 (2019).
57. T. Satyanarayana, S. Abraham, H. B. Kagan, *Angew. Chem. Int. Ed.* **48**, 456–494 (2009).
58. C. J. Wenthur *et al*., *J. Med. Chem.* **56**, 5208–5212 (2013).
59. A. Z. Gonzalez *et al*., *J. Med. Chem.* **57**, 2472–2488 (2014).
60. S. K. Gediya, G. J. Clarkson, M. Wills, *J. Org. Chem.* **85**, 11309–11330 (2020).
61. Q. An *et al*., *J. Am. Chem. Soc.* **145**, 359–376 (2023).
62. S. Kim, P.-P. Chen, K. N. Houk, R. R. Knowles, *J. Am. Chem. Soc.* **144**, 15488–15496 (2022).
63. S. Wilsey, P. Dowd, K. N. Houk, *J. Org. Chem.* **64**, 8801–8811 (1999).
64. A. Bensari, J.-L. Renaud, O. Riant, *Org. Lett.* **3**, 3863–3865 (2001).

ACKNOWLEDGMENTS

Funding: We thank the National Natural Science Foundation of China (grant nos. 22125111 and 21971163), National Key R&D Program of China (grant no. 2021YFA1500100), the Shanghai Pilot Program for Basic Research – Chinese Academy of Science, Shanghai Branch, and SIOC for financial support. We also thank J. Lipshultz for thoughtful discussions and J. Wu for assistance with the NMR experiments. **Author contributions:** L.W., J.D., and L.D. contributed equally to this work. Z.Z. conceived and directed the research; Z.Z., L.W., J.D., L.D., and S.W. designed the experiments. L.W., J.D., L.D., S.W., Q.A., and H.W. performed and analyzed the reactions; Z.Z., L.W., and S.W. prepared the manuscript, which was approved by all authors. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** Data are available in the manuscript and supplementary materials. The supplementary crystallographic data for this paper are available free of charge from the Cambridge Crystallographic Data Centre (CCDC) under accession numbers 2265923, 2265930, 2266206, and 2287759. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj0040
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S42
References (*65*–*69*)

Submitted 1 June 2023; accepted 4 September 2023
10.1126/science.adj0040

PLANT SCIENCE

A dirigent protein complex directs lignin polymerization and assembly of the root diffusion barrier

Yi-Qun Gao[1]†, Jin-Quan Huang[2]†, Guilhem Reyt[1]‡, Tao Song[2,3], Ashley Love[4], David Tiemessen[4], Pei-Ying Xue[1,5], Wen-Kai Wu[2,3], Michael W. George[4], Xiao-Ya Chen[2,3], Dai-Yin Chao[2]*, Gabriel Castrillo[1]*, David E. Salt[1]*

Functionally similar to the tight junctions present in animal guts, plant roots have evolved a lignified Casparian strip as an extracellular diffusion barrier in the endodermis to seal the root apoplast and maintain nutrient homeostasis. How this diffusion barrier is structured has been partially defined, but its lignin polymerization and assembly steps remain elusive. Here, we characterize a family of dirigent proteins (DPs) essential for both the localized polymerization of lignin required for Casparian strip biogenesis in the cell wall and for attachment of the strip to the plasma membrane to seal the apoplast. We reveal a Casparian strip lignification mechanism that requires cooperation between DPs and the Schengen pathway. Furthermore, we demonstrate that DPs directly mediate lignin polymerization as part of this mechanism.

All living organisms have evolved mechanisms to regulate their nutrient homeostasis by shielding their cells from external fluctuations. In vascular plants, root permeability is controlled by the endodermis, a specialized root tissue layer, along with its Casparian strip (CS) and is critical for plant survival in natural ecosystems (*1*). To achieve this, the CS seals the apoplastic space between neighboring endodermal cells using lignin-based cell wall impregnations (*2*). This sealing mechanism allows the endodermis to act as a selective gate for water and mineral nutrients to enter the vascular system while also preventing them from leaking out, mimicking the function of the tight junctions in animal epithelium that control gut permeability (*3*).

The activity of the plant NADPH oxidase respiratory burst oxidase homolog F (RBOHF) catalyzes the production of hydrogen peroxide (*4*, *5*), which is used as a substrate by the Casparian-strip–localized peroxidases involved in the generation of reactive monolignols required for lignin polymerization (*4*, *6*). ENHANCED SUBERIN1 (ESB1) (*7*), a dirigent protein (DP), is also required for lignin deposition at the CS. It localizes to the CS, though its biochemical role has not been elucidated. (*8*, *9*). DPs are known to control stereoselectivity in monolignol dimerization to free lignans (*10*–*13*), however their role in higher order lignin polymerization, also considering a random process, has been controversial for more than two decades (*14*). An additional role in surveillance controlling the sealing of the finished CS is played by the Schengen pathway (*15*–*17*). Nonetheless, until now no clear integrative mechanism has emerged for the polymerization of lignin in the CS.

[1]Future Food Beacon of Excellence & School of Biosciences, University of Nottingham, Sutton Bonington, UK. [2]National Key Laboratory of Plant Molecular Genetics, CAS Centre for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China. [3]University of Chinese Academy of Sciences, Beijing 100049, China. [4]School of Chemistry, University of Nottingham, Nottingham, UK. [5]State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory for Farmland Eco-environment of Hebei Province, College of Resources and Environmental Sciences, Hebei Agricultural University, Hebei, Baoding 071000, China.
***Corresponding author. Email: dychao@cemps.ac.cn (D.-Y.C.); gabriel.castrillo@nottingham.ac.uk (G.C.); david.salt@nottingham.ac.uk (D.E.S.)**
†These authors contributed equally to this work.
‡Present address: LIPME, Université de Toulouse, INRAE, CNRS, 31326 Castanet-Tolosan, France.

An endodermal family of DPs is associated with the Casparian strip

We examined the expression patterns of 20 DPs out of a total of 25 identified DPs in *Arabidopsis thaliana* using the available literature (*18*). We found that five DP genes, *DIR9*, *DIR24*, *DIR16*, *DIR18*, and *ESB1* showed a specific root endodermal expression pattern (fig. S1A) and were found to be down-regulated in the *myb36* mutant (fig. S1B), with a reduced activity of MYB36, a master regulator of the CS biosynthesis machinery (*19*). These data suggest that DIR9, DIR24, DIR16, and DIR18 together with ESB1 might have a function in lignin polymerization associated with the CS. For further analysis, we also included in our selection DIR25, a low-expressed, uncharacterized homolog of ESB1 that is also regulated by MYB36 (fig. S1B and S1C). These six proteins represent a phylogenetic cluster, classified as DIR-e (*20*), in *A. thaliana* (Fig. 1A).

Thus, we expressed the six DPs fused to the fluorescent protein mCherry to characterize their localization patterns. We confirmed the localization of all selected DPs at the endodermis, specifically at the CS (Fig. 1, B to E). Although all six endodermal DPs localized to the edges of the CS, ESB1 and DIR25 accumulated more in the central region of the CS, DIR9 and DIR24 localized to both the central region and the edges of the CS, and DIR16 and DIR18 mainly localized to the edges of the CS (Fig. 1, F and G, and fig. S2). The CS localization patterns of these proteins are in line with the presence of characteristic domains in their sequences (Fig. 1B and fig. S3). We therefore classified this family of DPs into three groups: group I (ESB1 and DIR25), group II (DIR9 and DIR24), and group III (DIR16 and DIR18) (Fig. 1B).

We found that DPs from group I and II are localized at the cell wall of the endodermal cells whereas the proteins from group III are present at both the cell wall and the cytoplasm of endodermal cells (fig. S4A). We noted that the pattern of localization of group III proteins resembles the one for the deposition of the polyester suberin in the endodermis (*21*, *22*), but we found that the localization pattern of group III DPs and endodermal suberin biosynthesis were not associated (fig. S4, B and C). Thus, we discarded the participation of the group III DPs in suberin biosynthesis.

Next, we studied whether the presence of characteristic domains in the DP structures determines their localization. In ESB1 (group I), DIR9 (group II), and DIR16 (group III) fused to mCherry, we deleted the endodermal (EN) domain, which is common to all the six endodermal DPs (fig. S5, A to C). The lack of EN domain did not affect DP localization at the CS, indicating that the EN domain alone is not sufficient to determine DP localization (fig. S5, A to C). In the ESB1-mCherry protein we deleted the glycine-rich domain (GRD), which is uniquely present in group I DPs (ESB1-ΔGRD-mCherry, fig. S5D). A high proportion of ESB1-ΔGRD-mCherry was then detected in the cytoplasm of endodermal cells compared with the intact ESB1 that mainly localized in the cell wall (fig. S5D). In DIR9 fused to mCherry we removed the uncharacterized domain (UN), present only in group II DPs, (DIR9-ΔUN-mCherry, fig. S5E). Similar to ESB1-ΔGRD-mCherry (fig. S5D), DIR9 without the UN domain localizes more in the cytoplasm of endodermal cells and to a lesser extent in the cell wall (fig. S5E). These findings resemble the localization pattern found in group III DPs in which the GRD and UN domains are not present (fig. S4A). Further, the insertion of the GRD domain of ESB1 (group I) into the sequence of DIR16 (group III), led DIR16 to fully localize to the cell wall (fig. S5F). Therefore, the GRD and UN domains, present in group I and II DPs, are necessary for the complete cell wall localization of these proteins at the CS.

DPs are required for Casparian strip integrity

We next created (*dir25*, *dir16*, and *dir24*) and identified [*esb1* (*7*), *dir9*, and *dir18*] knockout mutants for all six DP genes (fig. S6, A to E). In the single *dir* mutants, we examined lignin deposition at the CS using specific lignin staining with basic fuchsin and also quantified the

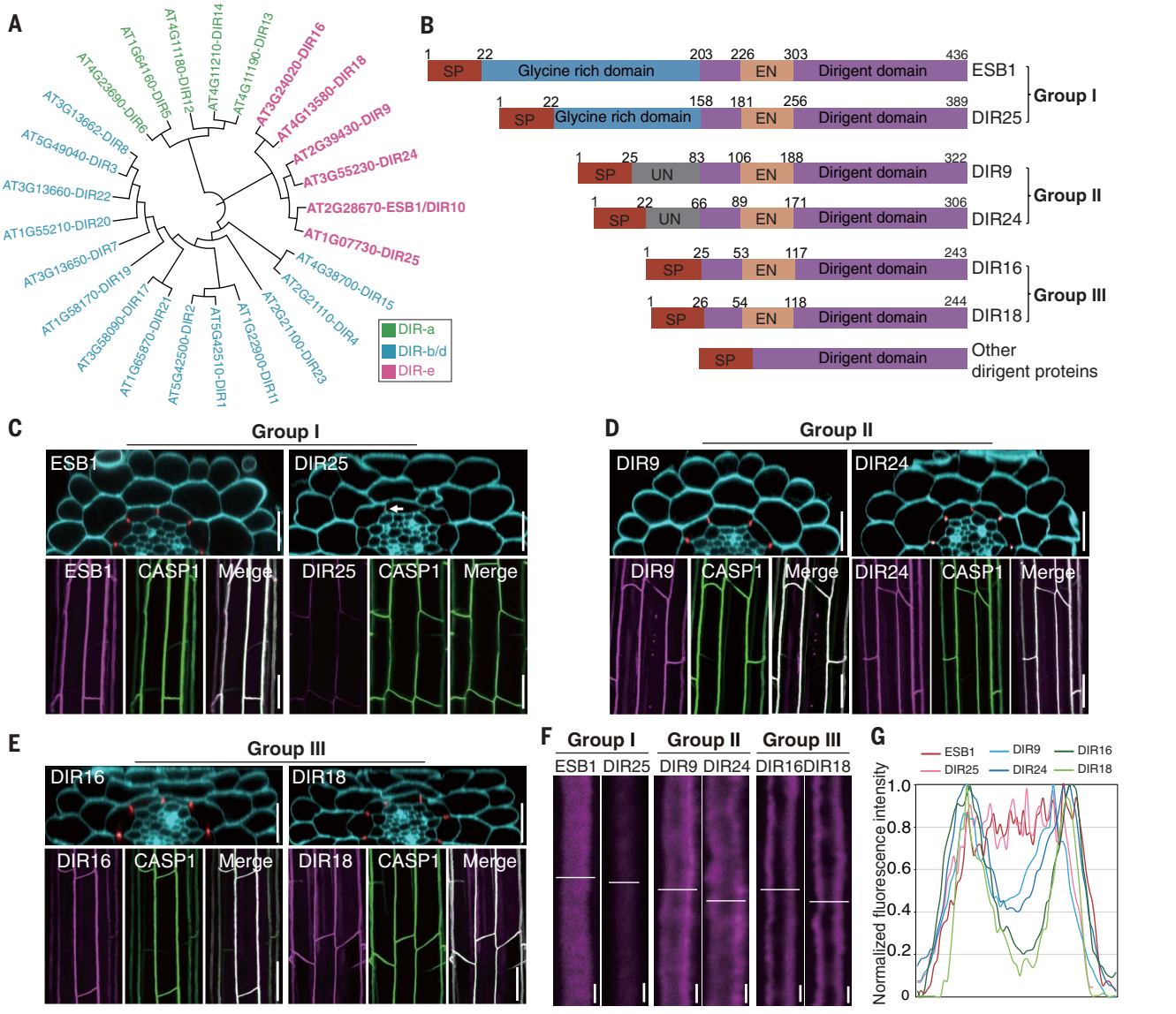


Fig. 1. The endodermal family of DPs is associated with the Casparian strip. (**A**) Phylogenetic tree built using the sequences of the 25 DPs identified in *A. thaliana*. Colors denote the different DP families. (**B**) Schematic diagrams representing the domains found in the protein sequences in the three groups of the endodermal family of DPs. The red regions represent signal peptides (SP). The numbers above each diagram are the position of the first amino acid in each domain. (**C** to **E**) Confocal microscopy images showing the localization of the DP family members in the root. The upper panels show root cross sections showing the endodermal expression at the CS of the DPs from group I (C), group II (D), and group III (E). The lower panels show colocalization of the DP family members with CASP1 (CASP1-GFP). The white arrow in the upper panel of (C) indicates the DIR25-mCherry expression. All pictures were taken at the 10th cell after the onset of elongation of endodermal cells. (**F**) Confocal microscopy images showing surface views of all DP family members localized at the CS. Horizontal white lines represent the trajectory used to quantify the pixel fluorescence intensity in (G). (**G**) Graph showing the distribution of normalized pixels fluorescence intensity values (normalized fluorescence intensity) along the CS used as a proxy to quantify the expression level of the DP family members. Scale bars are equivalent to 20 μm in the cross-sections (C, D, and E, upper panels), 10 μm in (C, D, and E, lower panels), and 1 μm in (F). We repeated this quantification 5 times.

diffusion of the cellular tracer propidium iodide (PI) into the vasculature. These two parameters provide information about the integrity of the CS as a diffusion barrier (fig. S6, F and G). We confirmed the defective CS lignin phenotype in the *esb1* mutant with high diffusion of PI (fig. S6G). By contrast, single mutants of *dir25*, *dir9*, *dir24*, *dir16*, and *dir18*, have normal CS lignin deposition (fig. S6F), and PI permeability, similar to wild-type (WT) plants (fig. S6G).

To further explore the role of the DPs in CS biosynthesis, we created high-order mutants among all family members and analyzed their CS integrity. In all the double mutants within groups [*esb1dir25* (group I), *dir9dir24* (group II), and *dir16dir18* (group III)] we observed phenotypes compatible with an impaired CS that are different from those found in the *esb1* mutant (Fig. 2, A and B). Phenotypes included gaps in the normally continuous lignin band, leaving "islands" of lignin, along with ectopic lignification at the endodermal cell corners (Fig. 2, A and B). These double *dir* mutants lost 50% of the lignified region of the CS when compared with WT plants (Fig. 2C), and all showed significantly increased PI penetration (Fig. 2D).

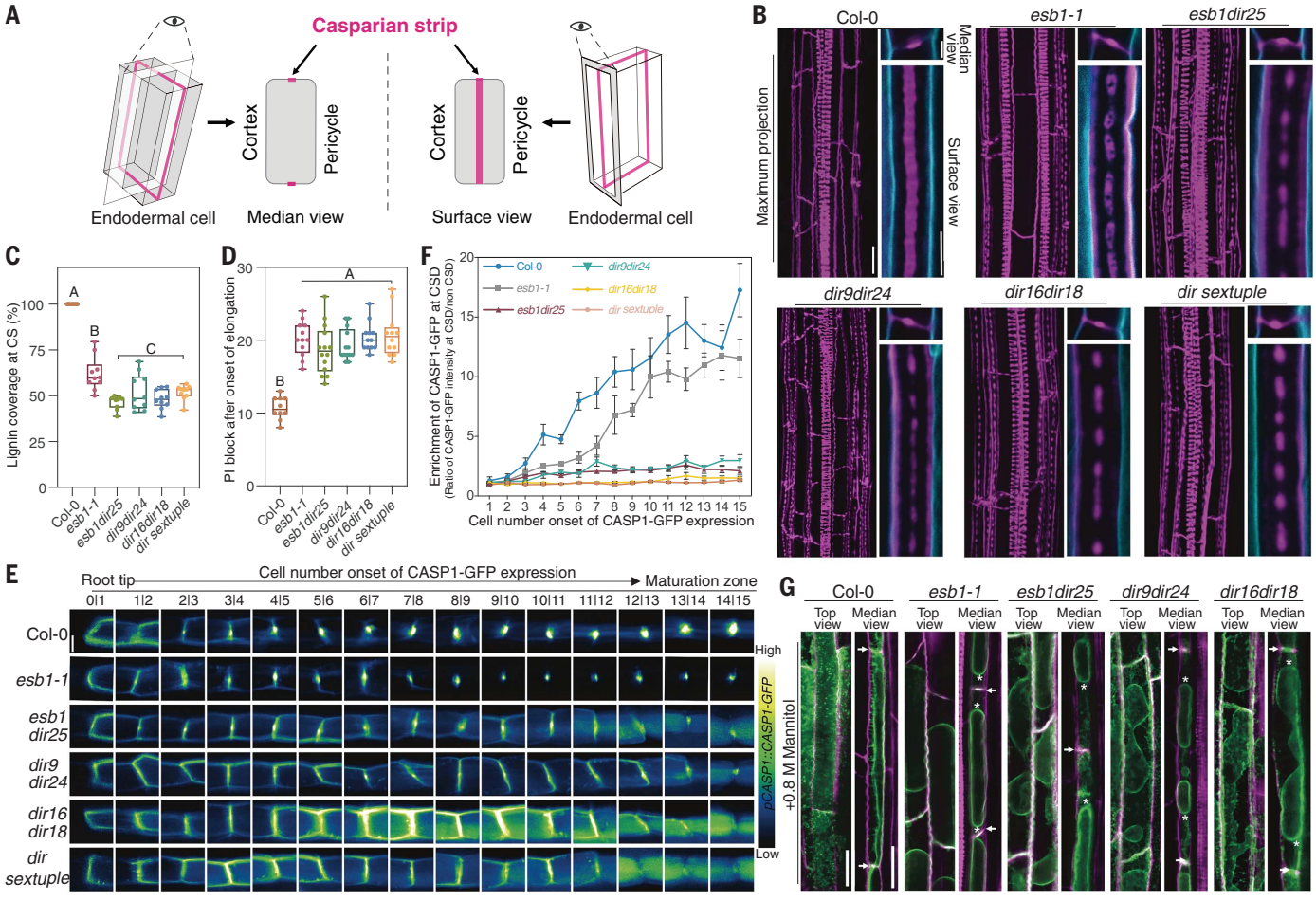


Fig. 2. The endodermal family of DPs is required for Casparian strip lignification, CASP1 localization at the CSD, and sealing of the endodermal apoplast. (**A**) Schematic diagram of an endodermal cell with the CS (magenta belt) showing the optical sections, median view, and surface view, used in (B). (**B**) Confocal images showing maximum projection, median view, and surface view of endodermal lignification in Col-0 plants, *esb1-1*, double DP mutants *esb1dir25*, *dir9dir24*, *dir16dir18*, and sextuple mutant *esb1dir25dir9dir24dir16dir18*. Lignin deposition (magenta) and cell walls (cyan) were visualized at the 15th cell after the onset of elongation. Scale bars are equivalent to 10 μm in the maximum projections, 5 μm in the surface views, and 2.5 μm in the median views. (**C**) Quantification of lignin coverage at the surface of the CS in WT Col-0 and DP mutant plants. Letters indicate significant difference [$P < 0.01$, one-way analysis of variance (ANOVA) with Tukey's honestly significant difference test], $n = 10$ to 14. (**D**) Quantification of propidium iodide diffusion in roots of Col-0 and the DP mutants used. Letters indicate significant difference ($P < 0.01$, one-way ANOVA with Tukey's honestly significant difference test), n =10. (**E**) Confocal images show CASP1-GFP (green) localization at the CSD. Scale bars, 2.5 μm. (**F**) Quantification of CASP1-GFP enrichment at the CSD from (E). Data represent mean ± SD, $n = 8$. (**G**) SYP122 (green) localization in the 10th endodermal cells after the onset of elongation in Col-0 and the double DP mutants. The arrows signal the adhesion region between the plasma membrane and the cell wall at the CSD. Asterisks show the fully detached plasma membranes. Scale bars, 20 μm.

Supporting the role of the DPs in lignin deposition in the CS, we found that the double mutants *esb1dir25*, *dir9dir24*, and *dir16dir18* have increased endodermal suberization (fig. S7), a compensatory mechanism observed in mutants with a compromised CS function. Further, we found a significantly different accumulation of mineral nutrients in the leaves of the double *dir* mutants compared with WT Col-0 plants, which is consistent with an impaired CS (fig. S8 and table S1). These results indicate that the endodermal family of DPs is required for lignin deposition in the CS, with consequences for mineral nutrient accumulation in the plant.

To assess the genetic relationship among different DP groups, we generated higher-order mutants including triple (*esb1dir16dir18* and *esb1dir9dir24*), quadruple (*dir16dir18dir9dir24*), quintuple (*esb1dir16dir18dir9dir24*), and sextuple (*esb1dir16dir18dir9dir24dir25*) mutants among endodermal DP group members. In all cases, we found that these combinatorial high-order mutants phenocopied the double mutants *esb1dir25*, *dir9dir24*, and *dir16dir18* (Fig. 2, B to D, figs. S7, S8, and S9, A to C). Therefore, we concluded that higher-order *dir* mutants, comprising all different DP group combinations, did not enhance the Casparian strip–related phenotypes observed in the double *dir* mutants. This shows that each group of DPs perform specific, nonredundant functions, working in the same pathway to ensure localized lignin deposition at the CS. One possibility is that these DPs act as part of the same multiprotein hetero-complex.

To test this hypothesis, we performed bimolecular fluorescence complementation (BiFC) experiments expressing combinatorial pairs of DPs in *Nicotiana benthamiana* leaves. We demonstrated that DPs from different groups can interact with each other and form heterodimers, for example: ESB1-DIR16, DIR9-DIR16, and DIR9-ESB1 (fig. S10A). Furthermore, using co-immunoprecipitation assays, we validated the

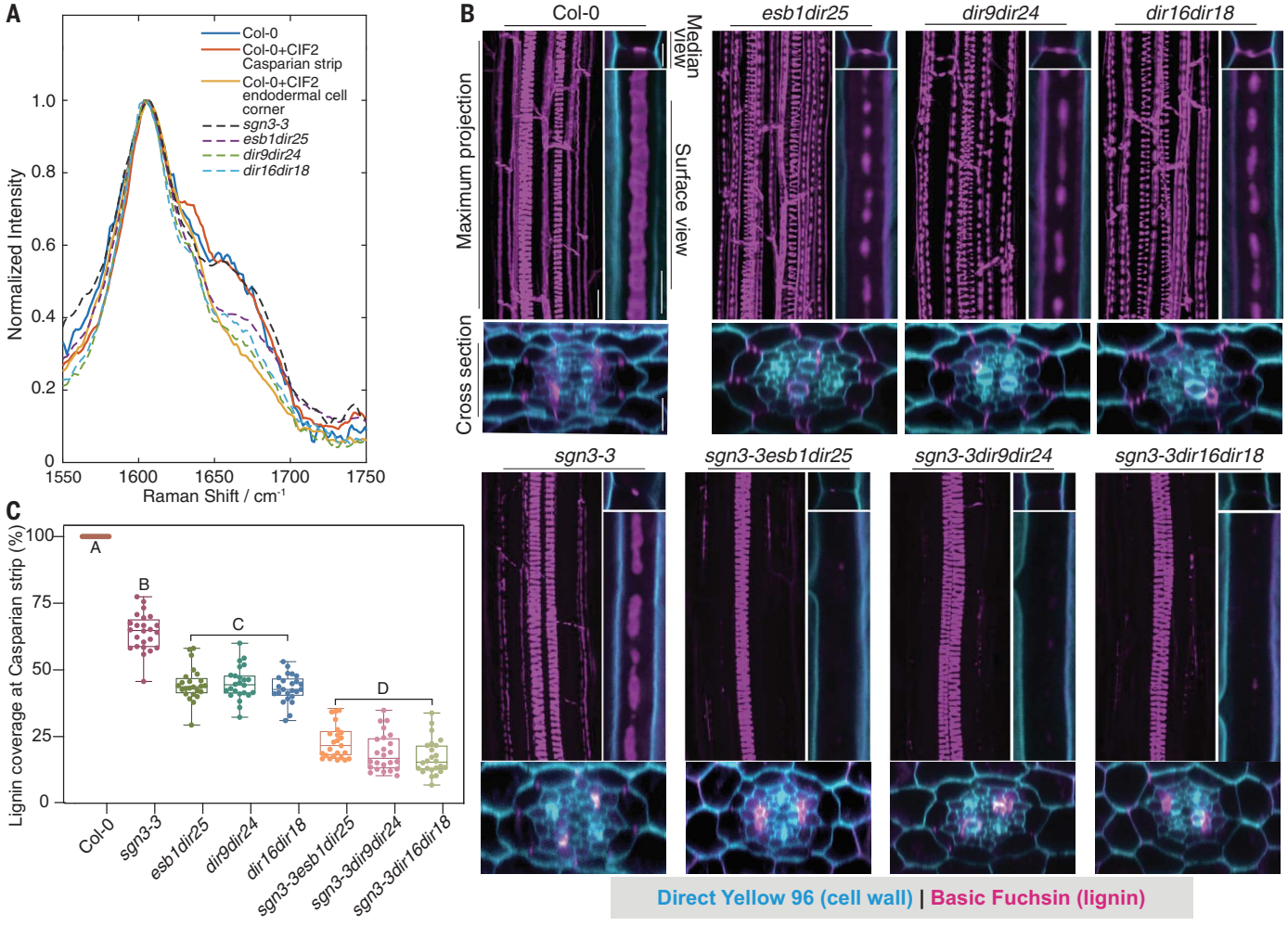


Fig. 3. Casparian strip lignification requires cooperation between the endodermal DPs and the Schengen pathway. (**A**) Representative Raman spectral components in the region corresponding to lignin aromatic rings (1550 to 1750 cm^{-1}) of CS lignin obtained from WT Col-0, *sgn3-3* and double DP mutant plants, *esb1dir25*, *dir9dir24*, and *dir16dir18*, obtained using multivariate curve resolution (MCR) analysis; $n = 3$. (**B**) Confocal images showing maximum projection, median views, surface views, and cross sections of endodermal lignin deposition in WT Col-0, *sgn3-3*, double dirigent mutants *esb1dir25*, *dir9dir24*, and *dir16dir18*, and triple mutants *sgn3-3esb1dir25*, *sgn3-3dir9dir24*, and *sgn3-3dir16dir18* plants. Lignin deposition and the cell walls were visualized at the 15th endodermal cell after the onset of elongation using basic fuchsin staining (magenta) and Direct Yellow 96 (cyan), respectively. Vertical white lines are multiple comparison scale bars equivalent to 10 μm in the maximum projection and cross section images and equivalent to 5 and 2.5 μm in the surface and median view images, respectively. (**C**) Graph showing the quantification of lignin coverage at the CS [lignin coverage at CS (percent)] in WT Col-0, double dirigent mutants, and triple *sgn3-3 dir* mutants. Different letters in the figure indicate significant difference ($P < 0.01$, one-way ANOVA with Tukey's honestly significant difference test), n =18 for Col-0, and $n = 24$ for the rest of the genotypes.

existence of DP heterocomplexes of higher stoichiometric order (fig. S10B), which confirmed that ESB1, DIR9, and DIR16—proteins from each group of endodermal DIRs—can form a heterotrimer.

To study whether CASP1—an integral plasma membrane protein involved in the localization and structuring of the CS (*23*)—is functionally related to endodermal DPs, we analyzed changes in the localization of CASP1-GFP at the CS domain (CSD), in *esb1-1* single mutant and all double, triple, quadruple, quintuple, and sextuple *dir* mutant backgrounds. The CSD is a specific region of the plasma membrane occupied by the CS biosynthetic machinery. We found that CASP1-GFP localization showed a clear disruption at the CSD in the higher order *dir* mutants (Fig. 2, E and F, and fig. S9D). We confirmed this finding in all CASP proteins (fig. S11). These results indicate that the endodermal family of DPs are necessary for the proper localization of CASPs at the CSD, to assemble a functional CS biosynthesis complex and initiate deposition of the CS.

We next evaluated the adhesion of the plasma membrane to the cell wall at the CSD. This is a characteristic of a functional CS. We expressed the plasma membrane protein SYP122 fused to mCitrine in the endodermis of WT, *esb1-1*, and *esb1dir25*, *dir9dir24*, and *dir16dir18* double mutant backgrounds to evaluate its localization at the CSD, as a probe for plasma membrane attachment to the CS. We found that SYP122-mCitrine diffused into the CSD in all double *dir* mutants (fig. S12), while we observed a typical exclusion zone for SYP122 at the CSD in WT plants, and a partial exclusion in the single mutant *esb1-1* (fig. S12). This result suggests that the attachment between the plasma membrane and the CS—required for exclusion of plasma membrane proteins such as SYP122—was fully interrupted in all double *dir* mutants and partially interrupted in *esb1-1*. We next validated this finding using mannitol-induced plasmolysis in *esb1-1* and the three double *dir* mutants (Fig. 2G). These results indicate that the endodermal DPs are necessary to anchor the endodermal plasma membrane

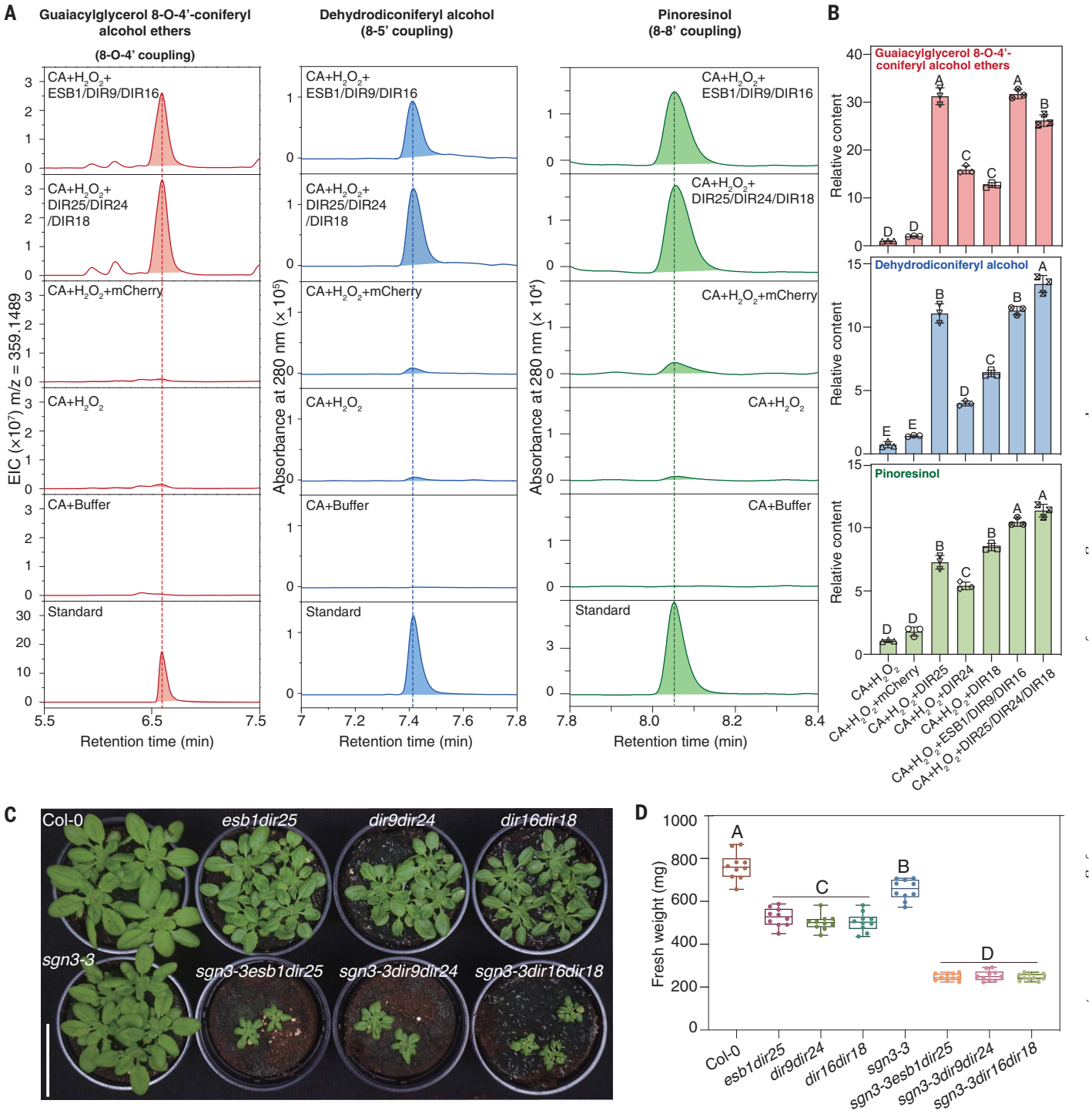


Fig. 4. Members of the endodermal family of DPs directly regulate Casparian strip monolignol polymerization. (**A**) Figure shows the UPLC-MS profiles of the products obtained in in vitro monolignol polymerization reactions. In the reactions we used coniferyl alcohol (CA), two hetero-complex mixtures of ESB1/DIR9/DIR16 and DIR25/DIR24/DIR18 supplemented with H_2O_2. As a control we used the same reaction without the presence of the DPs (CA + H_2O_2). Nonrelated mCherry protein was used as a negative control. On top of the figure are the three standards of lignan with different CA monomer coupling bond types used. (**B**) Quantification of new synthesized 8-O-4', 8-5' and 8-8' coupling lignan obtained in the in vitro monolignol polymerization reactions with and without the endodermal family of DPs present. Different letters in the graph indicate significant difference ($P < 0.01$, one-way ANOVA with Tukey's honestly significant difference test), $n = 3$. We repeated this experiment three times. (**C**) Pictures show the growth of 3-week-old plants of WT Col-0, double dirigent mutants *esb1dir25*, *dir9dir24*, *dir16dir18*, single mutant *sgn3-3*, and the triple mutants *sgn3-3esb1dir25*, *sgn3-3dir9dir24*, and *sgn3-3dir16dir18* in soil under a short-day regime. Scale bar, 5 cm. (**D**) The graph shows the fresh weight quantification of plants used in (C). The letters in the figure indicate significant difference ($P < 0.01$, one-way ANOVA with Tukey's honestly significant difference test), $n = 10$.

to the CS in the cell wall. Both CASP1 function and plasma membrane anchoring are independent of lignin deposition in the CS (*15*). Because CASP proteins are also necessary for plasma membrane attachment (*24*), we cannot exclude that DPs exert this function as part of a complex with CASPs.

Endodermal DPs are required for lignin synthesis at the Casparian strip

We studied the chemical nature of the remaining lignin islands observed in the double *dir* mutants *esb1dir25*, *dir9dir24*, and *dir16dir18*. We used confocal Raman microscopy on root cross sections to spectrally resolve the chemical features associated with the islands of lignin produced in these mutants (Fig. 3A and fig. S13). In all double mutants analyzed, we observed that the CS lignin spectrum was different from that found in WT plants (Fig. 3A and fig. S13). In line with a previous publication (*15*), the continuous WT CS lignin showed two characteristic Raman peaks: a lower intensity peak at 1604 cm^{-1} associated with an aromatic ring stretch (*25*–*27*), and a peak at 1658 cm^{-1} assigned to a vibrational mode with double bond (C=C and C=O) character conjugated to an aromatic ring (e.g., coniferyl alcohol or coniferaldehyde) present in the lignin (*25*–*28*). The islands of lignin present in the CS of double *dir* mutants predominantly exhibited the peak at 1604 cm^{-1}, and the peak at 1658 cm^{-1} was substantially smaller (Fig. 3A and fig. S13A). Further, a peak around 1333 cm^{-1}, assigned to sinapyl alcohol in lignin from an intact CS (*26*), was much lower in all *dir* mutants compared with that found in the wild type (fig. S13A). The Raman spectra extracted from the islands of lignin in *dir* double mutants were similar to those extracted from the lignin deposited in the endodermal cell corners of WT plants treated with the CASPARIAN STRIP INTEGRITY FACTOR 2 (CIF2) peptide (Fig. 3A and fig. S13) (*15*). This peptide is the ligand of the SCHENGEN3 (SGN3) receptor that activates lignin polymerization required to close gaps in the developing CS (*14*). This Schengen pathway also leads to ectopic deposition of lignin in the cell corners when the Casparian strip is physically impaired or CIF peptides are applied exogenously (*16*, *29*). Therefore, these results support the notion that the lignin islands observed in the double *dir* mutants are chemically distinct from the typical lignin found in the continuous CS of WT plants. It is likely then that the lignin islands remaining in the double *dir* mutants are synthesized by a different lignin biosynthetic complex that requires the Schengen pathway.

We next generated the triple mutants *sgn3esb1dir25*, *sgn3dir9dir24*, and *sgn3dir16dir18*. Consistent with our hypothesis, lignin deposition in the CS of these mutants was almost undetectable (Fig. 3B). Furthermore, the lignin coverage in the CS of these triple mutants was reduced by over 80%, compared with the 25% reduction observed in *sgn3-3* (Fig. 3C). These results confirmed that the endodermal family of DPs are required for the biosynthesis of most of the lignin at the CS with the remaining islands of lignin observed in the double *dir* mutants requiring the Schengen pathway. These lignin islands are not an artifact of hyperactivation of the Schengen pathway in the double *dir* mutants because they require MYB36 for their formation (*15*, *16*).

Development of the CS starts with lignin being deposited in discontinuous discrete formations, which has been termed the "string-of-pearls" stage (*20*). We found that the endodermal DPs colocalized with this initiation lignin in WT plants (fig. S14A). Further, the first appearance of CS lignin was delayed in the *esb1dir25*, *dir9dir24*, and *dir16dir18* double mutants compared with WT Col-0 plants, despite all genotypes having similar root lengths (fig. S14, B to E). We propose that in WT plants, as development progresses, the gaps between the string-of-pearls initiation lignin are filled in by a process requiring the Schengen pathway. Corroborating this we found no delay in formation of the string-of-pearl lignin in the mutant *sgn3-3* (fig. S14, B to E), despite the early expression of *SGN3* in the endodermis (fig. S14F), higher expression of endodermal DPs compared with SGN pathway components where CS formation starts (fig. S14G), and only a partial colocalization of SGN3 with the initiation lignin (fig. S14A). Thus, the lignin islands we observe in the *dir* double mutant likely represent Schengen pathway lignin—here called gap-filling lignin. Correspondingly, the discontinuous CS lignin previously observed in the *sgn3-3* mutant represents the lignin initially deposited at the string-of-pearl stage that requires DPs—here called initiation lignin. Indeed, *sgn3-3* CS lignin showed similar Raman spectral characteristics to WT Col-0 (Fig. 3A and fig. S13), explained as DP-dependent lignin being a major component of CS lignin (Fig. 3C).

DPs directly drive the polymerization of lignin monomers

Next, we explored whether the endodermal DPs can be directly involved in lignin monomer (monolignol) polymerization. First, using a molecular docking approach we elucidated monolignol binding positions within the predicted three-dimensional (3D)-conformational structure of ESB1, DIR9, and DIR16 proteins. In all cases, H-bond formation was predicted between monolignol and amino acids from the EN domain in DPs (fig. S15, A to C). This suggests that the endodermal DPs—specifically the EN domain—might be directly involved in lignin polymerization.

We then generated mutated versions of the ESB1, DIR9, and DIR16 proteins lacking the EN domain. Unlike the intact DPs, in all cases these mutated versions of the DPs were unable to complement either the defects in CS lignification or CASP1 localization found in the *esb1*, *dir9dir24*, and *dir16dir18* mutants (fig. S15, D to F). Therefore, we verified that the EN domain is necessary for the endodermal DPs function in lignin polymerization.

We then expressed and purified all six endodermal DP members individually or combined using baculovirus-based expression vectors in insect cells. We obtained three individual DPs, DIR25, DIR24, and DIR18 (fig. S16A), as well as the control mCherry protein (fig. S16B), and two DP hetero-complex mixtures, ESB1/DIR9/DIR16 and DIR25/DIR24/DIR18 (fig. S16, C and D). To validate the direct function of DPs in monolignol polymerization, we reconstituted an in vitro lignin polymerization assay that included the substrate coniferyl alcohol, the most abundant monolignol present in the CS lignin, H_2O_2 as the electron acceptor, and the individual DPs alone or DP hetero-complex mixtures. In all cases we detected high yield of specific new peaks corresponding to different monolignol polymerization products only when the DPs were added to the reaction (Fig. 4, A and B, and fig. S17). These de novo synthesized products were identified using the corresponding standards as the dimerized monolignols guaiacylglycerol 8-O-4'-coniferyl ethers, dehydrodiconiferyl alcohol, and pinoresinol (Fig. 4A). Other controls validated this finding (Fig. 4A and fig. S17). Consistent with our previous results (fig. S10), more dimerized monolignol was generally detected when the DIR heterocomplex mixture was added to the reaction compared with the peaks observed in reactions catalyzed by individual DPs, especially for pinoresinol (Fig. 4B). This further suggests that the DPs act as a complex during lignin polymerization at the CS. Taken together, these results support the conclusion that the endodermal DPs directly mediate lignin polymerization during CS biosynthesis. In the absence of these DPs, reactive monolignols produced by the CS-localized RBOHF/Peroxidase system will no longer be polymerized at the CS and would diffuse freely to the cell corners and polymerize spontaneously through the simple combinatorial radical coupling process (*30*), forming the ectopic lignin we and others have observed (*6*, *15*).

DP activity influences plant adaptation to environmental stresses

We found that under short-day conditions, *dir* double mutants *esb1dir25*, *dir9dir24*, and *dir16dir18* exhibited significant reduced vegetative growth (~65% shoot fresh weight of the wild type). This retarded growth was more obvious in the *sgn3dir* triple mutants with ~32% shoot fresh weight of WT plants (Fig. 4, C and D). When we exposed the *dir* double mutants and the *sgn3dir* triple mutants to salinity, osmotic stress, iron toxicity, low ambient

humidity in agar plates and soils, all *sgn3dir* triple mutants showed significantly reduced root and shoot growth (fig. S18, A to E). Further, the *sgn3dir* triple mutants showed a significant reduction in the shoot area similar to that observed in the double mutant *sgn3myb36* lacking a CS (*15*) (fig. S18, A to D and F). We observed similar responses in plants grown in soils exposed to low ambient humidity (fig. S18, G and H). These results demonstrate that CS lignification—driven by the endodermal DPs—is required for plant growth and plays a role in enabling plants to adapt to challenging environments.

Conclusions

Our results reveal a mechanism that explains precise lignification at the CS that requires cooperation between the MYB*36* regulated endodermal DPs function and the Schengen pathway (fig. S19). All of this knowledge can now be generalized to other lignified plant tissues where DP are widely expressed, and redundancy is also expected.

REFERENCES AND NOTES

1. N. Geldner, *Annu. Rev. Plant Biol.* **64**, 531–558 (2013).
2. S. Naseer *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10101–10106 (2012).
3. M. Barberon, N. Geldner, *Plant Physiol.* **166**, 528–537 (2014).
4. Y. Lee, M. C. Rubio, J. Alassimone, N. Geldner, *Cell* **153**, 402–412 (2013).
5. S. Fujita *et al*., *EMBO J.* **39**, e103894 (2020).
6. N. Rojas-Murcia *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 29166–29177 (2020).
7. I. Baxter *et al*., *PLOS Genet.* **5**, e1000492 (2009).
8. P. S. Hosmani *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14498–14503 (2013).
9. Y. Wang *et al*., *Nat. Commun.* **13**, 2222 (2022).
10. L. B. Davin *et al*., *Science* **275**, 362–367 (1997).
11. D. R. Gang *et al*., *Chem. Biol.* **6**, 143–151 (1999).
12. K. W. Kim *et al*., *J. Biol. Chem.* **287**, 33957–33972 (2012).
13. B. Pickel *et al*., *Angew. Chem. Int. Ed.* **49**, 202–204 (2010).
14. R. Hatfield, W. Vermerris, *Plant Physiol.* **126**, 1351–1357 (2001).
15. G. Reyt *et al*., *Nat. Commun.* **12**, 2320 (2021).
16. V. G. Doblas *et al*., *Science* **355**, 280–284 (2017).
17. A. Pfister *et al*., *eLife* **3**, e03115 (2014).
18. K. H. Ryu, L. Huang, H. M. Kang, J. Schiefelbein, *Plant Physiol.* **179**, 1444–1456 (2019).
19. T. Kamiya *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10533–10538 (2015).
20. S. Ralph, J. Y. Park, J. Bohlmann, S. D. Mansfield, *Plant Mol. Biol.* **60**, 21–40 (2006).
21. T. G. Andersen *et al*., *Nature* **555**, 529–533 (2018).
22. T. G. Andersen, M. Barberon, N. Geldner, *Curr. Opin. Plant Biol.* **28**, 9–15 (2015).
23. D. Roppolo *et al*., *Nature* **473**, 380–383 (2011).
24. I. C. R. Barbosa *et al*., *Nat. Commun.* **14**, 1626 (2023).
25. N. Gierlinger, M. Schwanninger, *Plant Physiol.* **140**, 1246–1254 (2006).
26. P. N. Perera, M. Schmidt, V. L. Chiang, P. J. Schuck, P. D. Adams, *Anal. Bioanal. Chem.* **402**, 983–987 (2012).
27. K. L. Larsen, S. Barsberg, *J. Phys. Chem. B* **114**, 8009–8021 (2010).
28. A. M. Saariaho, D. S. Argyropoulos, A. S. Jaaskelainen, T. Vuorinen, *Vib. Spectrosc.* **37**, 111–121 (2005).
29. T. Nakayama *et al*., *Science* **355**, 284–286 (2017).
30. J. Ralph *et al*., *Phytochem. Rev.* **3**, 29–60 (2004).

ACKNOWLEDGMENTS

We thank N. Geldner, University of Lausanne, for kindly sharing the *pCASP2/3/4/5::CASP2/3/4/5-GFP* and *pSGN3::SGN3-mVenus* expression plasmids and transgenic lines. We thank W. Yin and S-Y. Shi, University of Nottingham for sharing the BiFC vectors; the Nanoscale and Microscale Research Centre at the University of Nottingham for the Raman microscope facility; S.-S. Wang, L.-Y. Jing, X.-Y. Xu, and M.-L. Ma at the CAS Centre for Excellence in Molecular Plant Science for the help in the DP characterization and the mass spectra analysis. We also thanks E. Shani, Tel Aviv University, Israel, and T.G. Andersen, Max Planck Institute for Plant Breeding, Germany, for critical comments on the manuscript. **Funding:** This study was funded by the Royal Society Newton International Fellowship (NIF\R1\191915) to Y-Q.G. and D.E.S.; the Royal Society Newton Advanced Fellowship (NAF\R1\201264) to D-Y.C. and D.E.S.; UK Biotechnology and Biological Sciences Research Council Grant (grant BB/L027739/1) to D.E.S.; the European Research Area Network for Coordinating Action in Plant Sciences (grant ERACAPS13.089_RootBarriers) to D.E.S.; National Natural Science Foundation of China (31930024) and Chinese Academy of Sciences (XDB27010103) to D-Y.C.; the Future Food Beacon of Excellence at the University of Nottingham to Y-Q.G. and D.E.S.; the BBSRC-NSF grant (BB/V011294/1) to G.C.; the National Natural Science Foundation of China (32388201) to X-Y.C.; the National Key R&D Program of China (2022YFF1001400) to J-Q.H.; the Foundation of Youth Innovation Promotion Association of the Chinese Academy of Sciences to J-Q.H., the Young Elite Scientists Sponsorship Program by CAST (2019QNRC001) to J-Q.H., and the EPSRC grant EP/S017739/1 and the University of Nottingham Propulsion Futures Beacon of Excellence to the Raman microscope facility. **Author contributions:** Conceptualization: Y.-Q.G., J.-Q.H., D.-Y.C., G.C. and D.E.S. Methodology: Y.-Q.G., J.-Q.H., T.S., A.L., D.T., G.C. and D.E.S. Investigation: Y.-Q.G., J.-Q.H., G.R., T.S., A.L., D.T., P.-Y.X., W.-K.W. Visualization: Y.-Q.G., J.-Q.H., G.R., T.S., A.L., D.T. Funding acquisition: Y.-Q.G., D.E.S., G.C., D-Y.C., J.-Q.H., X.-Y.C. Project administration: G.C. and D.E.S. Supervision: G.C. and D.E.S. Writing – original draft: Y.-Q.G. and G.C. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data are available in the main text or supplementary materials. The plasmids, seeds of mutants and the transgenic *A. thaliana* plants used or generated in this study are available upon request. **License information:** Copyright 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi5032
Materials and Methods
Figs. S1 to S19
Tables S1 and S2
References (*31*–*46*)
MDAR Reproducibility Checklist

Submitted 30 April 2023; accepted 7 September 2023
10.1126/science.adi5032

WORKING LIFE

By **Michał Turek**

A lesson from failure

"Thank you so much for taking the time to apply for the position," the email read. "Your application has not been successful. We wish you every success in securing a suitable position in the future." I received the email 2 weeks after interviewing for a postdoc position. When I read it, I felt disappointed and angry—and not just because I hadn't gotten the job. I was disappointed because I'd spent hours preparing for the interview and in the end, I didn't even know why I wasn't successful. "How can I adapt my job search approach if I don't know what went wrong?" I thought. Now, when early-career scientists interview for positions in my lab, I keep that experience in mind. If they don't get the position, I want them to know why and what they can do to improve their applications going forward.

The value of that kind of feedback became obvious to me a few months after I received the postdoc rejection email. After applying for a job at a consulting company, I spent months working my way through the exhaustive recruiting and interview process before getting a phone call from a company representative. "Unfortunately, we decided to offer the job to someone else," the voice said. Again I felt deflated. But the conversation wasn't over. "Would you like to hear feedback from us?" they asked.

The representative went on to tell me I hadn't clearly articulated my motivation for applying for the position and why I wanted to work for the company. The feedback was tough to hear. But I quickly realized they were right. After completing my Ph.D. in Germany, my top priority had been to find a position in Poland, where I grew up. As a result, I wasn't very selective about what I applied for; I pursued any job opportunity that broadly fit my skills and interests. The job at the consulting company was something I was technically capable of doing, but I didn't feel passionate about it—and that clearly showed in the interview process.

From then on, I decided to carefully evaluate each job opening and only pursue ones that truly resonated with what I was passionate about. In my applications, I also began stating more clearly and emphatically why I wanted that particular job. Soon the approach started to pay off, as I got an offer for a postdoc position in Poland that I was truly excited about.

A few years (and several more successful interviews) later, I found myself on the other side of the interview table.

"If no one tells us where we are lagging, we are liable to make the same mistakes over and over again."

As a freshly hired group leader, I was interviewing candidates for the first Ph.D. position in my lab. I was particularly looking forward to talking with one candidate; on paper, his application looked great—nice CV, good recommendations, strong cover letter. But the interview went badly. He didn't seem to have done any reading in advance about the research I was doing, which made me wonder whether he was truly interested in working with me.

When I called to tell him I couldn't offer him the position, I could tell, even over the phone, that he was very disappointed. So I asked him if he wanted to know where he could improve in future interviews. It's never pleasant to listen to—or dole out—constructive criticism. But as the conversation went on, I got the sense that he was grateful for what I had to say.

I've since filled several positions in my lab. For each recruitment, I've made a point to treat every interview as a trade: Candidates invest time and effort to prepare for our conversation—and if they don't get the position, the least I can do is reward their effort with pointers about how they can improve. And when my own trainees apply for positions elsewhere, I plan to encourage them to ask for feedback at the end of the interview process. It's something I wish I had done myself earlier in my career. Because if no one tells us where we are lagging, we are liable to make the same mistakes over and over again. ■

Michał Turek is a group leader at the Institute of Biochemistry and Biophysics of the Polish Academy of Sciences. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.